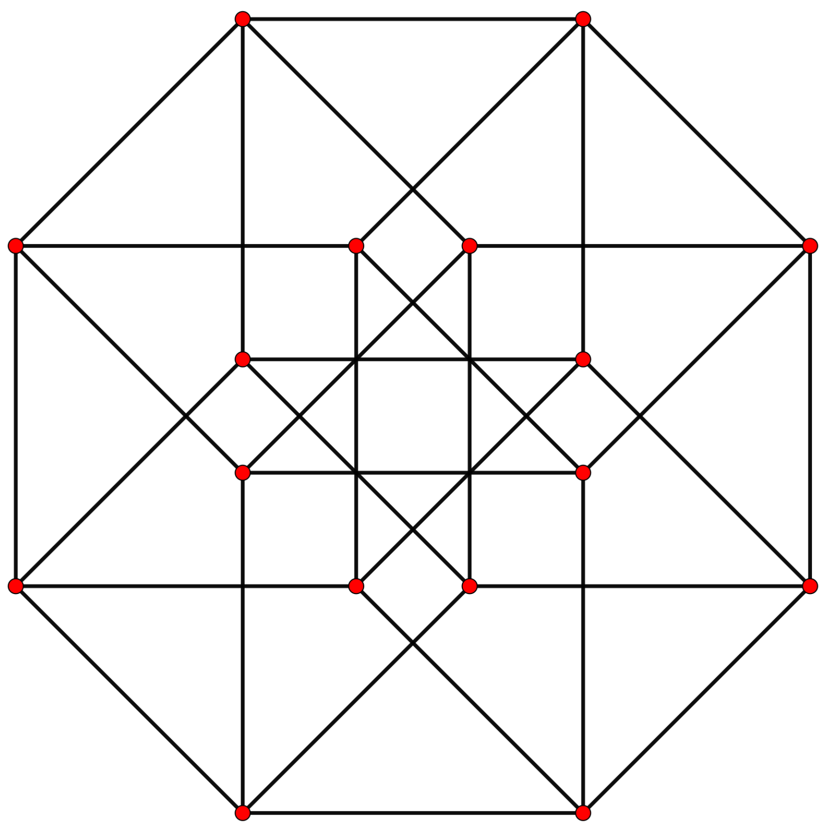
MATH 115
Linear Algebra for Engineering
Fall 2025

Faculty of Mathematics
University of Waterloo



$$H = \left\{ \vec{x} \in \mathbb{R}^4 \;\middle|\; -1 \leq \vec{x} \cdot \vec{e}_i \leq 1,\ i = 1, 2, 3, 4 \right\}$$

For Yang and Kevin.

Acknowledgements

These course notes have slowly evolved from a simple collection of typeset lecture notes that were created in the spring term of 2020. I am grateful to Steven Furino, who encouraged me to create these notes in place of a standard textbook, and I give my thanks to Logan Crew, Ghazal Geshnizjani, Matthew Harris, Aaron Hutchinson, Carrie Knoll and Michelle Molino, each of whom generously contributed to those lecture notes, greatly improving their accuracy, readability and clarity.

In 2023, I was given the necessary support to update my lecture notes into this set of course notes. In alphabetical order, I am grateful to:

- Faisal Al-Faisal for the time he has dedicated to helping me put these course notes together, for providing me with many great comments and suggestions on how to further improve their readability, and for editing the many practice problems throughout the notes.
- Eddie Dupont for reading through these course notes and raising several issues with me that made me rethink how to best present the material in this course. Eddie also spent a considerable amount of time creating the Python exercises which illustrate the usefulness of linear algebra as well as the importance of using a programming language to handle large-scale linear algebra problems that students will surely encounter in the real world.
- Jordan Hamilton who additionally reviewed the course notes and assisted me in the rather large task of coordinating MATH 115, affording me the breathing room to continue modifying and correcting this document.

I consider myself very fortunate to have had the opportunity to teach linear algebra alongside Keith Nicholson at the University of Calgary and Dan Wolczuk at the University of Waterloo. Both of these professors have shown the utmost dedication to their students which has greatly influenced how I teach today. This set of course notes is inspired heavily by the outstanding linear algebra textbooks they have each written in the past.

Thank you to my students, both past and present, for always asking so many great questions which force me to think about linear algebra concepts in a new way. After teaching an introductory course in linear algebra many times, I am still always surprised by how much I continue to learn from students each time I teach this course

Of course, I cannot forget to say thank you to my lovely wife, Yang Zhou, who has had to endure my working late on countless nights over the last few years as a result of my both coordinating MATH 115 and eventually creating these course notes. Words can't convey how grateful I am for her unending patience and support, and I might add, for making sure I always stayed fed and watered.

Finally, my thanks go to Michael A. La Croix for creating and sharing their LaTeX style file which has led to a more readable (and colourful) set of MATH 115 course notes.

Contents

0 Introduction **7**
- 0.1 About these Course Notes . . . 7
- 0.2 Tips For Success in MATH 115 (and your other courses) . . . 10

1 Vector Geometry **13**
- 1.1 Vectors in $\mathbb{R}^n$ . . . 13
- 1.2 Linear Combinations . . . 24
- 1.3 The Norm and the Dot Product . . . 29
- 1.4 Vector Equations of Lines and Planes . . . 43
- 1.5 The Cross Product in $\mathbb{R}^3$ . . . 51
- 1.6 The Scalar Equation of Planes in $\mathbb{R}^3$ . . . 55
- 1.7 Projections . . . 60
  - 1.7.1 Shortest Distances . . . 62
- 1.8 Optional Section: Area and Volume . . . 68

2 Systems of Linear Equations **73**
- 2.1 Introduction and Terminology . . . 73
- 2.2 Solving Systems of Linear Equations . . . 79
- 2.3 Rank . . . 90
- 2.4 Homogeneous Systems of Linear Equations . . . 98
- 2.5 Comments on Combining Elementary Row Operations . . . 102

3 Matrices **105**
- 3.1 Matrix Algebra . . . 105
- 3.2 The Matrix–Vector Product . . . 112
- 3.3 The Matrix Equation $A\vec{x} = \vec{b}$ . . . 120
- 3.4 Matrix Multiplication . . . 127
- 3.5 Matrix Inverses . . . 134
  - 3.5.1 Matrix Inversion Algorithm . . . 137
  - 3.5.2 Properties of Matrix Inverses . . . 139

4 Subspaces of $\mathbb{R}^n$ **143**
- 4.1 Spanning Sets . . . 143
- 4.2 Geometry of Spanning Sets . . . 152
- 4.3 Linear Dependence and Linear Independence . . . 164
  - 4.3.1 Definition and Basic Results . . . 164
  - 4.3.2 Removing Dependencies . . . 171
  - 4.3.3 Further Examples . . . 177
- 4.4 Subspaces of $\mathbb{R}^n$ . . . 181
- 4.5 Bases and Dimension . . . 187

4.5.1 Bases of Subspaces . . . 187
4.5.2 Dimension of a Subspace . . . 195
4.5.3 Extending a Basis . . . 198
4.6 Fundamental Subspaces Associated with a Matrix . . . 202
4.7 Summary . . . 209

5 Linear Transformations **211**
5.1 Matrix Transformations and Linear Transformations . . . 211
5.2 Examples of Linear Transformations . . . 222
5.3 Operations on Linear Transformations . . . 232
5.4 Inverses of Linear Transformations . . . 242
5.5 The Kernel and the Range . . . 247

6 Determinants **255**
6.1 Determinants and Invertibility . . . 255
6.2 Elementary Row and Column Operations . . . 264
6.3 Properties of Determinants . . . 272
6.4 Optional Section: Area, Volume and Determinants . . . 278
6.5 Optional Section: Adjugates and Matrix Inverses . . . 287

7 Complex Numbers **295**
7.1 Basic Operations . . . 295
7.2 Conjugate and Modulus . . . 302
7.3 Polar Form . . . 308
7.3.1 Powers of Complex Numbers . . . 312
7.3.2 Complex Exponential Form . . . 313
7.4 Complex Polynomials . . . 317
7.5 Complex nth Roots . . . 324

8 Eigenvalues and Eigenvectors **329**
8.1 Introduction . . . 330
8.1.1 Example: Reflections Through Lines in $\mathbb{R}^2$ . . . 330
8.1.2 Example: Projections onto Planes in $\mathbb{R}^3$ . . . 335
8.2 Computing Eigenvalues and Eigenvectors . . . 340
8.3 Eigenspaces . . . 348
8.4 Diagonalization . . . 355
8.5 Powers of Matrices . . . 364

A A Brief Introduction to Sets **369**

B Solutions to Exercises **375**

Chapter 0

Introduction

0.1 About these Course Notes

The material in these course notes may be presented in a way that you are not entirely familiar with from high school. You will likely find the content in these notes (as well as the material presented in the lectures) to be more terse and move at a faster pace than what you have experienced before. Although perhaps daunting at first, students are, by and large, able to adapt to the faster pace of university within the first couple of weeks.

It can help, in this course at least, to understand how these course notes present the material. As you can see from the table of contents, there are 8 chapters (not counting this one), each containing multiple sections (with some sections having subsections). Aside from the narrative in each section that strives to add further explanations and put the material being learned into the context of what has been previously taught, these notes can be thought of as consisting of four "parts": definitions, examples, theorems and exercises. We briefly explain the importance of each of these, starting with definitions.

Definition 0.1.1

Key words from the definition will appear here for easy reference

There is a lot of new language introduced in linear algebra, and definitions are how we will present this new vocabulary to you. Key words in the definition will always appear in **boldface** so a quick glance will tell you what the definition is about.

This definition is called Definition 0.1.1. The first number refers to the Chapter number (this is Chapter 0), the second number refers to the section (this is Section 0.1), and the third number refers to this being the first definition in this section. A similar convention is used for examples and theorems.

The importance of definitions *cannot* be underestimated. Many mathematicians will agree that they are in fact the most important part of learning mathematics! For example, consider the following:

> Let $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^3$. Show that if $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent, then $\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$ is linearly independent.

As someone who is only beginning to learn linear algebra, you likely have no idea how to approach this problem. This will largely be due to the fact that you probably do not know what any of

$$\vec{v}_1, \quad \vec{v}_2, \quad \vec{v}_3, \quad \in, \quad \mathbb{R}^3, \quad \text{linearly independent} \quad \text{or} \quad \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$$

mean (and that is completely okay). These will all be presented throughout the course as definitions. Once you have understood these definitions, you will have a better understanding of how to approach the above problem. It is not uncommon for a student to perform poorly on an assessment question simply because they did not know the definitions required to understand the problem.

Example 0.1.2 **(Examples appear in green boxes)**

Examples are the main way we will present how to *do* linear algebra, and the course notes are filled with them. Many are straightforward, designed to ensure your understanding of a definition or to help motivate a theorem. Others are more involved, sometimes revealing unexpected connections between different areas of linear algebra. Others act as counterexamples to show that a seemingly correct hypothesis is actually false.

Success in linear algebra hinges on your ability to both understand the material presented in an example and to emulate the illustrated methods in related problems. Although reading through an example can give you some insight into the workings of linear algebra, you will gain a deeper understanding of the content if you can solve problems presented in the examples on your own (without peeking at the solutions!).

Theorem 0.1.3 **(Theorems appear in blue boxes)**

The many important results of linear algebra are presented as theorems. It is through these theorems that you will see the how mathematicians take existing results and definitions and combine them to to logically create new results. Some theorems state basic properties about a recently defined operation, while others present very deep (and sometimes surprising or unexpected) results.

Many of the theorems you will encounter in linear algebra will be followed by a proof of the theorem, particularly when the proof helps develop your insight into how linear algebra "works". Indeed, many of the proofs of theorems in linear algebra are concise, straightforward, elegant and instructive. [1]

In MATH 115, you will be expected to provide proofs of some basic results. As an engineer, you might ask how you would ever benefit from doing this. The short answer, which you will probably find less than satisfying, is that it's good for you. The longer answer offers several reasons: by learning how to writing proofs, you will

- learn how derive formulas rather than simply use them,

[1] There are some exceptions, of course - many of theorems presented in Chapter 6 are omitted because they can be computationally difficult and do not offer any additional insight beyond what you would gain simply by understanding the statement of the theorem.

- become efficient at manipulating expressions rather than simply plugging quantities into them to obtain an answer,
- be able to effectively decide if a given statement is true or false, and develop the necessary evidence to support your claim,
- be better able to present complex ideas and concepts to your colleagues as well as your employers, and
- learn how to generalize the results learned in this course and apply them to many areas of engineering - for example, machine learning is a current “hot topic” and relies heavily on many results from linear algebra, so much so that machine learning drives a lot of the current research in the field of linear algebra.

As a result of spending the time required to write proofs, you will gain a better understanding of the connections between the various topics of linear algebra, leading to you obtaining a deeper level of knowledge. You will begin to achieve a greater appreciation of the mathematics you are learning and you will find it easier to remember and recall the many concepts we cover in this course - a skill which will certainly be useful when you start using linear algebra in your later courses and future careers.

Exercise 0 **(In-Section Exercises appear in red boxes)**

It’s best to think of these exercises as checkpoints where you can verify if you have understood the material up to that point. Solving them will go a long way to ensure you don’t have any misunderstandings before moving forward with the content. Solutions to the exercises appear in Appendix B. Note that the exercises are numbered differently than the definitions, examples and theorems.

In addition to the in-section exercises, each section is followed by a series of practice problems (referred to as the End-of-Section Problems) which focus on both the computational and theoretical aspects of linear algebra. It is *highly* recommended that you attempt these problems and seek assistance if you are struggling with them. Solutions for the end-of-section problems appear in the accompanying solutions file.

0.2 Tips For Success in MATH 115 (and your other courses)

What follows is a list of things you can do to help increase your chances for success in MATH 115. This list is by no means exhaustive, and you will find that you constantly discover new habits such as these that will help you further succeed in your courses as you progress through your university career, and equally importantly, you will find certain habits that are detrimental to your success. It is important that you can distinguish between these habits and eliminate the ones that are not benefiting you. As you will see, part of university is figuring out what works for you, and what doesn't work. The best time to start doing this is now.

- **Eat, sleep and exercise**: These are the three most important things you can do to maintain your physical and mental health, but they will often be the first things to get cut from your schedules when you get busy. When you make your weekly schedules, be sure to include time for eating three meals per day, time for sleep, and time for exercise. If you are well-fed, well-rested and exercise regularly, you will find that you are more productive when it comes time to study and work on your assignments.

- **Start preparing for your assessments early:** It is never a good idea to begin an assignment the day it is due or to start studying for a quiz or test the night before your write it. Starting an assignment the day it is due will leave you with little time to understand the problems, think creatively about them, develop solutions, write coherent responses, or even finish all of the problems on time. By only preparing for a quiz or a tutorial the night before, you rob yourself of the time that is required to synthesize what you have learned in the lectures as well as the time needed to attempt multiple practice problems and discover which topics you are struggling with. Many quizzes and tutorial assignments have a time-limit, so you will need be efficient when solving problems, and this efficiency won't be achieved through last-minute studying. Starting to study just before a timed assessment can also lead to increased stress if you discover that you don't understand the material as well as you thought you did.

- **Vary your schedule**: Aside from eating, sleeping and exercise, your schedule should obviously have time set aside to work on each course as well as any assignments. It's tempting to create a schedule where each subject has a particular day, for example, you study calculus on Monday, linear algebra on Tuesday, etc. This is not the most effective way to study as the brain can only stay focused on one subject for so long. Instead, aim to include time each day to work on each of your courses.

- **Take frequent breaks**: Try to avoid working for more than an hour before taking a break. The longer you work without a break, the less productive you will become. If you find yourself surfing the web or watching videos on YouTube when you should be working, it's probably time to get up and stretch for a few minutes and maybe have a snack. When you return to work, you will likely find that your focus has returned.

- **Practice**: The more work you put into MATH 115, the more you will get out of it. In this course, you will be introduced to concepts that seem strange and abstract when first encountered. With a little hard work, you can begin to master these concepts and start to make important connections between the different topics covered throughout the semester. The end-of-section problems are designed to help you better understand the material presented during the lectures and it is highly recommended that you attempt them and ask questions if you are struggling with any of them.

- **Ask for help**: If you don't understand a concept, at least half of the students don't understand the concept! Never be afraid or ashamed to ask a question. Your instructor is here to help and is happy to do so. You can reach them
  - during office hours: see the Course Outline in the Course Information folder on LEARN for a listing of your instructor's office hours,
  - by email: see the Course Outline in the Course Information folder on LEARN for their email address.

  Additionally, you may speak with
  - the Engineering Instructional Support Tutors (EISTs), who will be a valuable resource that you are encouraged to take advantage of. Their hours of availability will also appear (once they are available) in the Course Information folder on LEARN,
  - Your classmates, with whom you are encouraged to discuss the course material.
- **Review your graded work**: Many students simply receive their graded assessments, look at the score and then don't think about it again. However, learning from your mistakes is one of the best ways to increase your knowledge! If you made an error on a question, try to understand why your solution to that question was not correct so that you don't make the same mistake again. If you received full marks for a question, then compare your answer to the posted solutions - perhaps the posted solution uses a different approach that will give you some new insight into the problem.
- **Have fun**: Engineering students typically have rather hectic schedules - which makes it even more important to schedule a bit of time away from all of your responsibilities each week! Try to have a day each week where you do something you enjoy that is not school related. We understand that you may not be able to do this every week, but you will feel recharged after taking some time away from school. There are also plenty of clubs at the University of Waterloo that you can join if you are looking to meet new people!

Chapter 1

Vector Geometry

1.1 Vectors in $\mathbb{R}^n$

We begin with the *Cartesian Plane*. We choose an origin O on this plane, and two perpendicular axes: a horizontal axis called the x_1-axis and a vertical axis called the x_2-axis.[1] We place these axes in the Cartesian plane so that they intersect at O. A point P in this plane is then represented by the ordered pair (p_1, p_2). If $p_1 = 0$, then P lies on the x_2-axis. Otherwise, we think of p_1 as a measure of how far to the right (if $p_1 > 0$) or how far to the left (if $p_1 < 0$) P is from the x_2-axis. If $p_2 = 0$, then P lies on the x_1-axis, otherwise we think of p_2 as a measure of how far above (if $p_2 > 0$) or how far below (if $p_2 < 0$) the x_1-axis P is. It then follows that the origin is the point $O(0, 0)$.

As we will see, it is often convenient to associate to each point a *vector* which we may view geometrically as an "arrow" or a directed line segment. Thus, given a point $P(p_1, p_2)$ in the Cartesian plane, we associate to it the vector

$$\vec{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}.$$

This is illustrated in Figure 1.1.1.

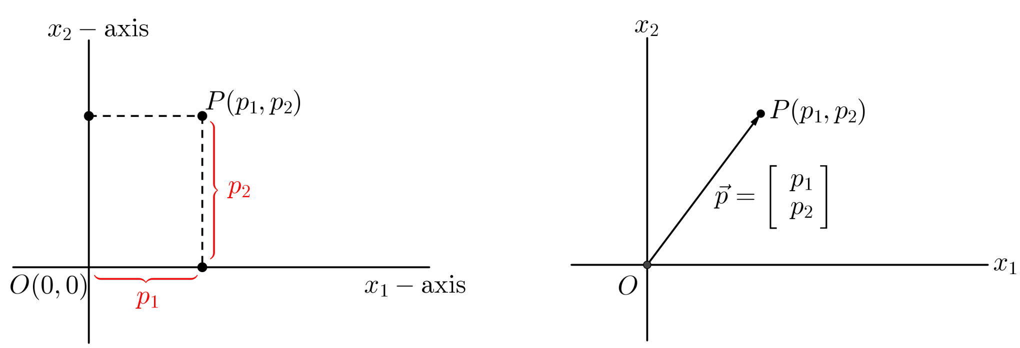


Figure 1.1.1: The point $P(p_1, p_2)$ in the Cartesian Plane and the vector $\vec{p} = \left[\begin{smallmatrix} p_1 \\ p_2 \end{smallmatrix}\right]$.

Of course, this idea extends to three-space where we have the x_1-, x_2- and x_3-axes as demonstrated in Figure 1.1.2.

[1] You might be more familiar with the names x-axis and y-axis. However, this naming scheme will lead to us running out of letters as we consider more axes, and hence we will call them the x_1-axis and the x_2-axis.

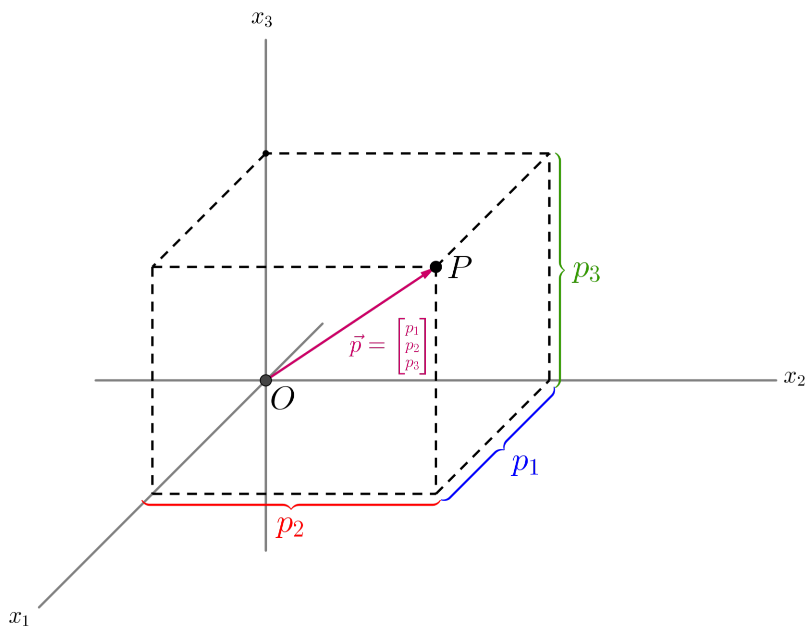


Figure 1.1.2: The point $P(p_1, p_2, p_3)$ and the vector $\vec{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$ in three-space. Note the order in which the three coordinate axes are labeled.

Definition 1.1.1

Vector, Component, $\mathbb{R}^n$

A **vector** $\vec{x}$ with n components is defined to be a column of n real numbers:

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{where } x_1, \ldots, x_n \in \mathbb{R}.$$

The numbers $x_1, \ldots, x_n$ are called the **components** (or **entries**) of $\vec{x}$.

The set of all vectors with n components is denoted by $\mathbb{R}^n$:

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \middle| \, x_1, \ldots, x_n \in \mathbb{R} \right\}.$$ [2]

In particular, we have

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| \, x_1, x_2 \in \mathbb{R} \right\} \quad \text{and} \quad \mathbb{R}^3 = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \middle| \, x_1, x_2, x_3 \in \mathbb{R} \right\}.$$

Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ are illustrated in Figure 1.1.1 and Figure 1.1.2, respectively.

Definition 1.1.2

Zero Vector

The **zero vector** in $\mathbb{R}^n$ is denoted by $\vec{0}_{\mathbb{R}^n} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$, that is, the vector whose n entries are all zero.

[2] Here we are using *set builder notation*. If this is unfamiliar to you, refer to Appendix A.

For example,

$$\vec{0}_{\mathbb{R}^2} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \vec{0}_{\mathbb{R}^3} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{0}_{\mathbb{R}^4} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and so on.}$$

We often simply denote the zero vector in $\mathbb{R}^n$ by $\vec{0}$ whenever this doesn't cause confusion. However, if we are considering, say, $\mathbb{R}^2$ and $\mathbb{R}^3$ at the same time, then we may prefer to write $\vec{0}_{\mathbb{R}^2}$ and $\vec{0}_{\mathbb{R}^3}$ to denote the zero vectors of $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively, since it may not be clear which zero vector we are referring to when we write $\vec{0}$.

Definition 1.1.3
Equality of Vectors

Two vectors $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$ are **equal** if $x_1 = y_1$, $x_2 = y_2$, $\dots$, $x_n = y_n$, that is, if their corresponding entries are equal. In this case, we write $\vec{x} = \vec{y}$ in this case. Otherwise, we write $\vec{x} \neq \vec{y}$.

Exercise 1

Is $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ equal to $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$?

It is important to note that if $\vec{x} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$ with $n \neq m$, then $\vec{x}$ and $\vec{y}$ can never be equal. For example, $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right] \neq \left[\begin{smallmatrix} 1 \\ 2 \\ 0 \end{smallmatrix}\right]$ as one vector belongs to $\mathbb{R}^2$ and the other belongs to $\mathbb{R}^3$.

We now begin to look at the algebraic operations that can be performed on vectors in $\mathbb{R}^n$. We will see that many of these operations are analogous to operations performed on real numbers and have very nice geometric interpretations.

Definition 1.1.4
Vector Addition

Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ be two vectors in $\mathbb{R}^n$. We define **vector addition** as

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \in \mathbb{R}^n,$$

that is, we add vectors by adding the corresponding entries.

Example 1.1.5

We have

- $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$.
- $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix}$.
- $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is not defined, because one vector is in $\mathbb{R}^3$ and the other is in $\mathbb{R}^2$.

We have a nice geometric interpretation of vector addition that is illustrated in Figure 1.1.3. We see that two vectors determine a parallelogram with their sum appearing as a diagonal of this parallelogram.[3]

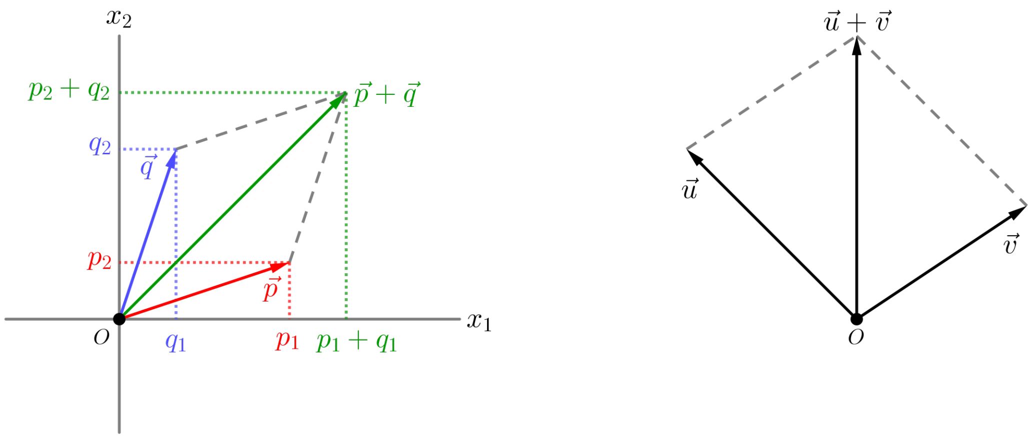


Figure 1.1.3: Geometrically interpreting vector addition. The figure on the left is in $\mathbb{R}^2$ with vector components labelled on the corresponding axes and the figure on the right is vector addition viewed for vectors in $\mathbb{R}^n$ with the x_1-, x_2-, $\ldots$, x_n-axes removed.

Definition 1.1.6

Scalar Multiplication

Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ and let $c \in \mathbb{R}$. We define **scalar multiplication** as

$$c\vec{x} = \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix} \in \mathbb{R}^n$$

that is, we multiply each entry of $\vec{x}$ by c. We call c a **scalar**, and say that $c\vec{x}$ is a **scalar multiple** of $\vec{x}$.

Example 1.1.7

We have

- $2\begin{bmatrix} 1 \\ 6 \\ -4 \\ 8 \end{bmatrix} = \begin{bmatrix} 2 \\ 12 \\ -8 \\ 16 \end{bmatrix}$.
- $0\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \vec{0}$.

[3] If the one of the two vectors being added is a scalar multiple of the other, then our parallelogram is simply a line segment or a "degenerate" parallelogram.

Figure 1.1.4 helps us understand geometrically what scalar multiplication of a nonzero vector $\vec{x} \in \mathbb{R}^2$ looks like. The picture is similar for $\vec{x} \in \mathbb{R}^n$.

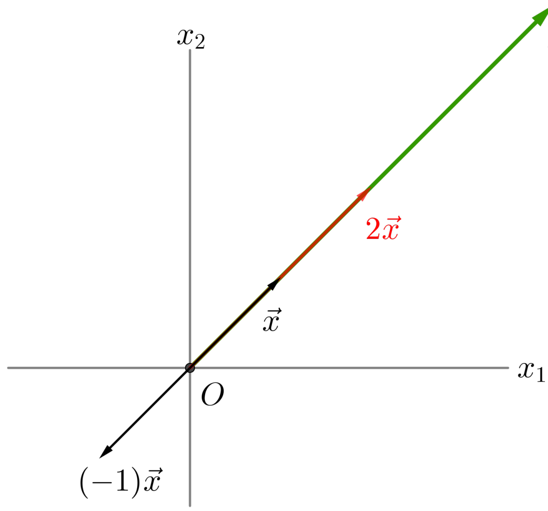


Figure 1.1.4: Geometrically interpreting scalar multiplication in $\mathbb{R}^2$.

Using the definitions of addition and scalar multiplication, we can define subtraction for $\vec{x}, \vec{y} \in \mathbb{R}^n$.

Definition 1.1.8
Vector Subtraction

Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. We define **vector subtraction** as

$$\vec{x} - \vec{y} = \vec{x} + (-1)\vec{y}.$$

Explicitly, if $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, then

$$\vec{x} - \vec{y} = \begin{bmatrix} x_1 - y_1 \\ \vdots \\ x_n - y_n \end{bmatrix},$$

that is, we subtract vectors by subtracting the corresponding entries.

Example 1.1.9

We have

- $\begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.
- $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 5 \end{bmatrix}$.
- $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is not defined, because one vector is in $\mathbb{R}^3$ and the other is in $\mathbb{R}^2$.

For $\vec{x}, \vec{y} \in \mathbb{R}^n$, we may think of the vector $\vec{x} - \vec{y}$ as the sum of the vectors $\vec{x}$ and $-\vec{y}$. This is illustrated in Figure 1.1.5. The picture is again similar in $\mathbb{R}^n$.

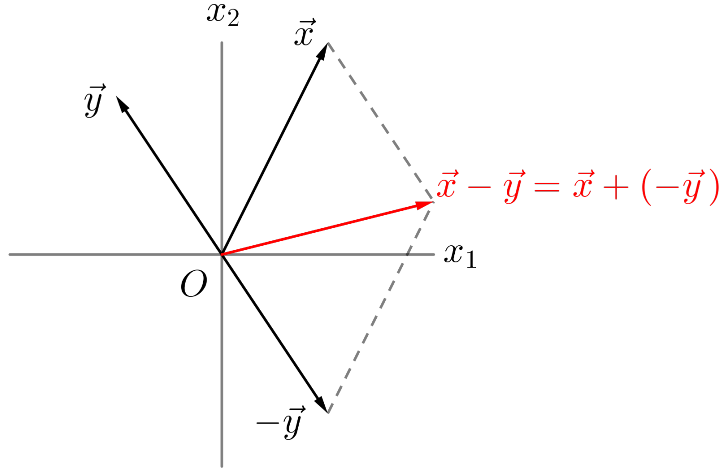


Figure 1.1.5: Geometrically interpreting vector subtraction in $\mathbb{R}^2$.

Exercise 2 Let $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}$. Determine a vector $\vec{z} \in \mathbb{R}^3$ such that

$$\vec{x} - 2\vec{z} = 3\vec{y}.$$

Thus far, we have associated vectors in $\mathbb{R}^n$ with points. Recall that given a point $P(p_1, \ldots, p_n)$, we associate with it the vector

$$\vec{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} \in \mathbb{R}^n$$

and view $\vec{p}$ as a directed line segment from the origin to P. Before we continue, we briefly mention that vectors may also be thought of as directed segments between arbitrary points. For example, given two points A and B in the x_1x_2-plane, we denote the directed line segment from A to B by $\overrightarrow{AB}$. In this sense, the vector $\vec{p}$ from the origin O to the point P can be denoted as $\vec{p} = \overrightarrow{OP}$. This is illustrated in Figure 1.1.6.

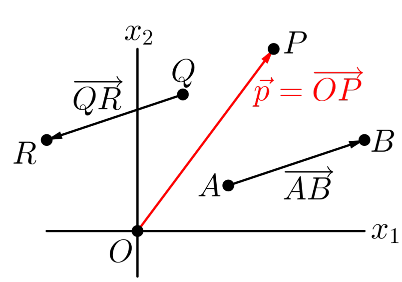


Figure 1.1.6: Vectors between points in $\mathbb{R}^2$.

Notice that Figure 1.1.6 is in $\mathbb{R}^2$, but that we can view directed segments between vectors in $\mathbb{R}^n$ in a similar way. We realize that there is something special about directed segments from the origin to a point P. In particular, given a point P, the entries in the vector $\vec{p} = \overrightarrow{OP}$ are simply the coordinates of the point P (refer to Figure 1.1.1 and Figure 1.1.2). Thus we refer to a vector $\vec{p} = \overrightarrow{OP}$ to be the *position vector* of P and and we say that $\vec{p}$ is in *standard position*. Note that in Figure 1.1.6, only the vector $\vec{p}$ is in standard position.

Finding a vector from a point A to a point B in $\mathbb{R}^n$ is also not difficult. For two points $A(a_1, a_2)$ and $B(b_1, b_2)$ we have that

$$\overrightarrow{AB} = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} - \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \overrightarrow{OB} - \overrightarrow{OA}$$

which is illustrated in Figure 1.1.7.

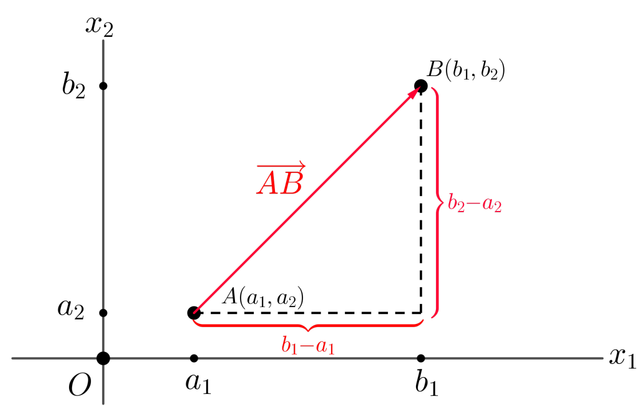


Figure 1.1.7: Finding the components of $\overrightarrow{AB} \in \mathbb{R}^2$.

This generalizes naturally to $\mathbb{R}^n$ where for $A(a_1, \ldots, a_n)$ and $B(b_1, \ldots, b_n)$ we have

$$\overrightarrow{AB} = \begin{bmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} - \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \overrightarrow{OB} - \overrightarrow{OA}.$$

Example 1.1.10

Find the vector from $A(1, 1, 1)$ to $B(2, 3, 4)$.

Solution: The vector from A to B is the vector $\overrightarrow{AB}$. We have

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

When we view vectors in $\mathbb{R}^n$ as directed segments between two points, our notation has a meaningful interpretation with regards to addition: given three points A, B and C, we have that

$$\overrightarrow{AC} = \overrightarrow{OC} - \overrightarrow{OA} = (\overrightarrow{OB} - \overrightarrow{OA}) + (\overrightarrow{OC} - \overrightarrow{OB}) = \overrightarrow{AB} + \overrightarrow{BC}.$$

Loosely speaking, travelling from A to C can be achieved from travelling first from A to B and then from B to C. This is illustrated in Figure 1.1.8.

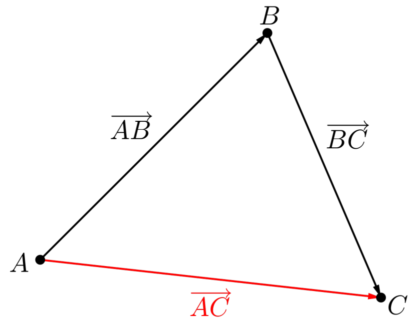


Figure 1.1.8: $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$.

Finally, putting everything together, we see that given two points A and B, their corresponding position vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ determine a parallelogram, and that the sum and difference of these vectors determine the diagonals of this parallelogram. This is displayed in Figure Figure 1.1.9, where the image on the right is obtained from the one on the left by setting $\vec{x} = \overrightarrow{OB}$ and $\vec{y} = \overrightarrow{OA}$. Note that by orienting vectors this way, $\overrightarrow{OB} - \overrightarrow{OA} = \vec{x} - \vec{y}$ is not in standard position.

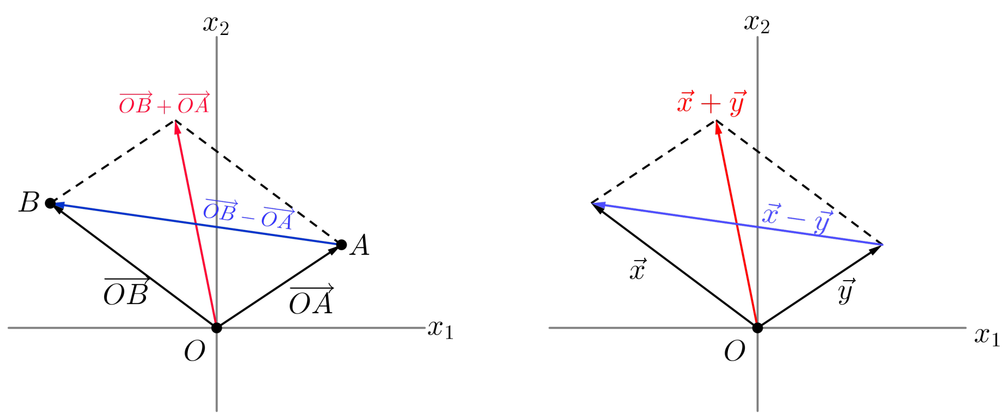


Figure 1.1.9: The parallelogram determined by two vectors. The diagonals of the parallelogram are represented by the sum and difference of the two vectors.

Having equipped the set $\mathbb{R}^n$ with vector addition and scalar multiplication, we state here a theorem that lists the properties these operations obey.

Theorem 1.1.11 **(Fundamental Properties of Vector Algebra)**

Let $\vec{w}, \vec{x}, \vec{y} \in \mathbb{R}^n$ and let $c, d \in \mathbb{R}$. We have

V1. $\vec{x} + \vec{y} \in \mathbb{R}^n$ — $\mathbb{R}^n$ is closed under addition

V2. $\vec{x} + \vec{y} = \vec{y} + \vec{x}$ — addition is commutative

V3. $(\vec{x} + \vec{y}) + \vec{w} = \vec{x} + (\vec{y} + \vec{w})$ — addition is associative

V4. $c\vec{x} \in \mathbb{R}^n$ — $\mathbb{R}^n$ is closed under scalar multiplication

V5. $c(d\vec{x}) = (cd)\vec{x}$ — scalar multiplication is associative

V6. $(c+d)\vec{x} = c\vec{x} + d\vec{x}$ — distributive law

V7. $c(\vec{x} + \vec{y}) = c\vec{x} + c\vec{y}$ — distributive law

These properties show that under the operations of vector addition and scalar multiplication, vectors in $\mathbb{R}^n$ follow very familiar rules. As we proceed through the course, we will begin to encounter some new algebraic objects and define operations on these objects in such a way that not all of these rules are followed.

Section 1.1 Problems

1.1.1 Let
$$\vec{x} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}.$$

Compute the following :

(a) $\vec{x} + \vec{y}$.
(b) $\vec{x} - \vec{y}$.
(c) $-2\vec{x}$.
(d) $3(\vec{x} + \vec{y})$.
(e) $2(3\vec{x} - \vec{y}) - 3(\vec{y} + 2\vec{x})$.

1.1.2 Consider the points $A(2, -1, -1)$, $B(3, 2, 4)$, and $C(1, 3, -2)$ in $\mathbb{R}^3$.

(a) Compute $\overrightarrow{AB}$.
(b) Show that $\overrightarrow{AB} = \overrightarrow{AC} + \overrightarrow{CB}$.
(c) Show that $\overrightarrow{AB} = \overrightarrow{AX} + \overrightarrow{XB}$ for any point X in $\mathbb{R}^3$.
(d) Show that $\overrightarrow{AB} = \overrightarrow{AX} + \overrightarrow{XY} + \overrightarrow{YB}$ for any points X, Y in $\mathbb{R}^3$.

1.1.3 Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and consider the statement

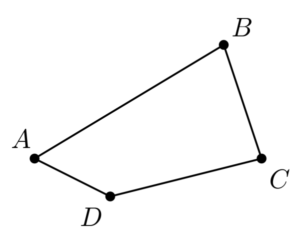
"If $\vec{x}$ is a scalar multiple of $\vec{y}$ then $\vec{y}$ is a scalar multiple of $\vec{x}$."

Either show this statement is true, or give an example that shows it is false.

1.1.4 Let $c \in \mathbb{R}$ and let
$$\vec{x} = \begin{bmatrix} 1 \\ 2 \\ c \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} c \\ c-1 \\ 1 \end{bmatrix}.$$

Determine all values of c so that $\vec{x}$ is a scalar multiple of $\vec{y}$.

1.1.5 Consider a quadrilateral $ABCD$ in $\mathbb{R}^3$ with vertices A, B, C, and D (as the figure below shows, the "name" $ABCD$ implies that edges of the quadrilateral are the segments AB, BC, CD and DA).



(a) Show that if $\overrightarrow{AB} = \overrightarrow{DC}$, then $ABCD$ is a parallelogram. [**Hint**: verify that $\overrightarrow{BC} = \overrightarrow{AD}$ which shows that opposite sides of $ABCD$ are parallel and of the same length.]
(b) Determine if the quadrilateral $ABCD$ with vertices $A(1, 2, 3)$, $B(2, -1, 4)$, $C(4, 3, 2)$, and $D(-1, -3, 5)$ is a parallelogram.
(c) Determine if the quadrilateral $PQRS$ with vertices $P(1, 4, -3)$, $Q(2, 5, 3)$, $R(-2, 3, 2)$ and $S(-3, 2, -4)$ is a parallelogram.

1.1.6 Let

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad \vec{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

be vectors in $\mathbb{R}^n$ and let $c, d \in \mathbb{R}$. Verify the following properties from the Fundamental Properties of Vector Algebra.

(a) V2.

(b) V3.

(c) V6.

1.1.7 Define a *computation* to be either the multiplication of two real numbers or the addition of two real numbers, and recall the Fundamental Properties of Vector Algebra.

(a) Property V5 states that $c(d\vec{x}) = (cd)\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$ and all $c, d \in \mathbb{R}$. Which of $c(d\vec{x})$ and $(cd)\vec{x}$ requires fewer computations to evaluate?

(b) Property V6 states that $(c+d)\vec{x} = c\vec{x} + d\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$ and all $c, d \in \mathbb{R}$. Which of $(c+d)\vec{x}$ and $c\vec{x} + d\vec{x}$ requires fewer computations to evaluate?

(c) Property V7 states that $c(\vec{x} + \vec{y}) = c\vec{x} + c\vec{y}$ for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and all $c \in \mathbb{R}$. Which of $c(\vec{x} + \vec{y})$ and $c\vec{x} + c\vec{y}$ requires fewer computations to evaluate?

1.2 Linear Combinations

In the previous section we learned about the two fundamental algebraic operations in linear algebra: vector addition and scalar multiplication. We will be frequently applying these operations to several vectors and scalars at the same time. For instance, every vector $\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$ in $\mathbb{R}^2$ can be obtained by scaling and adding the vectors $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

This motivates the following definition.

Definition 1.2.1 **Linear Combination**

Let $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k \in \mathbb{R}^n$ and $c_1, c_2, \dots, c_k \in \mathbb{R}$ for some positive integer k. We call the vector

$$c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k$$

a **linear combination** of the vectors $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$.

It follows from properties V1 and V4 of the Fundamental Properties of Vector Algebra that if we have $\vec{x}_1, \dots, \vec{x}_k \in \mathbb{R}^n$ and $c_1, \dots, c_k \in \mathbb{R}$, then the linear combination $c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k$ is also in $\mathbb{R}^n$. Thus every linear combination of $\vec{x}_1, \dots, \vec{x}_k$ will again be a vector in $\mathbb{R}^n$ and we say that $\mathbb{R}^n$ is *closed under linear combinations*.

Example 1.2.2

Evaluate the linear combination $4\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 5\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} - 4\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$.

Solution: We have

$$4\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 5\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} - 4\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 12 \end{bmatrix} + \begin{bmatrix} 5 \\ -10 \\ 5 \end{bmatrix} - \begin{bmatrix} 8 \\ 16 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ -18 \\ 13 \end{bmatrix}.$$

Example 1.2.3

Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ be such that $2\vec{x} - 5\vec{y} + 4\vec{z} = \vec{0}$. Express each of $\vec{x}, \vec{y}, \vec{z}$ as linear combinations of the other two vectors.

Solution: Solving the equation $2\vec{x} - 5\vec{y} + 4\vec{z} = \vec{0}$ for each of $\vec{x}, \vec{y}, \vec{z}$ gives

$$\vec{x} = \frac{5}{2}\vec{y} - 2\vec{z}, \quad \vec{y} = \frac{2}{5}\vec{x} + \frac{4}{5}\vec{z} \quad \text{and} \quad \vec{z} = -\frac{1}{2}\vec{x} + \frac{5}{4}\vec{y}.$$

Example 1.2.4

In $\mathbb{R}^3$, let

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

(a) Express $\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$ as a linear combination of $\vec{e}_1, \vec{e}_2, \vec{e}_3$.

(b) Express $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$ as a linear combination of $\vec{e}_1, \vec{e}_2, \vec{e}_3$.

Solution:

(a) For $c_1, c_2, c_3 \in \mathbb{R}$, consider

$$\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}.$$

Equating entries gives $c_1 = 1$, $c_2 = -2$ and $c_3 = 3$ so

$$\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = 1\vec{e}_1 - 2\vec{e}_2 + 3\vec{e}_3.$$

(b) Using the same method as in the first part, we have $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3$. This means that every $\vec{x} \in \mathbb{R}^3$ can be expressed as a linear combination of $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$.

Example 1.2.5

Let

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

If possible,

(a) express $\vec{u} = \begin{bmatrix} -1 \\ 2 \\ 7 \end{bmatrix}$ as a linear combination of $\vec{x}$ and $\vec{y}$.

(b) express $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$ as a linear combination of $\vec{x}$ and $\vec{y}$,

Solution:

(a) We want to find $c_1, c_2 \in \mathbb{R}$ such that

$$\begin{bmatrix} -1 \\ 2 \\ 7 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 \\ 2c_1 + c_2 \\ c_1 - c_2 \end{bmatrix}.$$

Equating entries gives the system of equations

$$\begin{array}{rcrcr} c_1 & + & c_2 & = & -1 \\ 2c_1 & + & c_2 & = & 2 \\ c_1 & - & c_2 & = & 7 \end{array}$$

Subtracting the first equation of this system from the second gives $c_1 = 3$ and it then follows from the first equation that $c_2 = -1 - c_1 = -1 - 3 = -4$. Since $c_1 - c_2 = 3 - (-4) = 7$, the third equation is also satisfied. Thus

$$\begin{bmatrix} -1 \\ 2 \\ 7 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 4 \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

(b) Proceeding in a similar manner, we want to find $c_1, c_2 \in \mathbb{R}$ such that

$$\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 \\ 2c_1 + c_2 \\ c_1 - c_2 \end{bmatrix}$$

which leads to the system of equations

$$\begin{array}{rcrcr} c_1 & + & c_2 & = & 1 \\ 2c_1 & + & c_2 & = & 2 \\ c_1 & - & c_2 & = & -4 \end{array}$$

As before, we subtract the first equation of this system from the second to obtain $c_1 = 1$ and it then follows from the first equation that $c_2 = 1 - c_1 = 1 - 1 = 0$. However, $c_1 - c_2 = 1 - 0 = 1 \neq -4$, so the third equation is not satisfied. Thus, $\vec{v}$ cannot be expressed as a linear combination of $\vec{x}$ and $\vec{y}$.

From Example 1.2.2, we see that it is straightforward to evaluate a linear combination. However, Example 1.2.5 shows that checking if a vector can be expressed as a linear combination of a collection of given vectors is more complicated. Such a problem involves solving a *system of equations* which can become tedious, even for few equations and few variables. The next chapter will investigate a more systematic approach to solving such systems of equations.

Exercise 3

(a) Show that $\left[\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right]$ is a linear combination of $\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$.

(b) Show that $\left[\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right]$ is not a linear combination of $\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 \\ -2 \end{smallmatrix}\right]$.

Section 1.2 Problems

1.2.1 Let

$$\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

If possible,

(a) express $\vec{x}$ as a linear combination of $\vec{v}_1$ and $\vec{v}_2$,

(b) express $\vec{y}$ as a linear combination of $\vec{v}_1$ and $\vec{v}_2$,

(c) express $\vec{x}$ as a linear combination of $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$.

1.2.2 Let

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \vec{z} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{u} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

(a) Express $\vec{u}$ as a linear combination of $\vec{x}$, $\vec{y}$ and $\vec{z}$ in 3 different ways.

(b) Find all coefficients $c_1, c_2, c_3 \in \mathbb{R}$ so that

$$\vec{u} = c_1\vec{x} + c_2\vec{y} + c_3\vec{z}.$$

[**Hint**: Write out the resulting system of equations (as done in Example 1.2.5) and realize that two of coefficients can be expressed in terms of the other coefficient.]

1.2.3 Let $\vec{x}, \vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^n$ and assume that $\vec{x}$ can be expressed as a linear combination of $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Show that if $\vec{v}_3$ can be expressed as a linear combination of $\vec{v}_1, \vec{v}_2$, then $\vec{x}$ can be expressed as a linear combination of just $\vec{v}_1$ and $\vec{v}_2$

1.2.4 Consider k arbitrary vectors $\vec{x}_1, \ldots, \vec{x}_k \in \mathbb{R}^n$.

(a) Show that the zero vector of $\mathbb{R}^n$ can be expressed as a linear combination of $\vec{x}_1, \ldots, \vec{x}_k$.

(b) Show that $\vec{x}_i$ can be expressed as a linear combination of $\vec{x}_1, \ldots, \vec{x}_k$ for each $i = 1, \ldots, k$.

1.2.5 Consider $\vec{x}, \vec{y}, \vec{z}, \vec{w} \in \mathbb{R}^n$. For each of the following, either show the statement is true, or give an example that shows it is false.

(a) If $\vec{x}$ can be expressed as a linear combination of $\vec{y}$ and $\vec{z}$, then $\vec{x}$ can be expressed as a linear combination of $\vec{y}$, $\vec{z}$, and $\vec{w}$.

(b) If $\vec{x}$ can be expressed as a linear combination of $\vec{y}$, $\vec{z}$ and $\vec{w}$, then $\vec{x}$ can be expressed as a linear combination of $\vec{y}$ and $\vec{z}$.

1.2.6 Consider two distinct nonzero vectors $\vec{x}, \vec{y} \in \mathbb{R}^2$. If possible,

(a) give an example of $\vec{x}, \vec{y}$ such that each of them can be expressed as a linear combination of the other.

(b) give an example of $\vec{x}, \vec{y}$ such that the conditions

- $\vec{x}$ can be expressed as a linear combination of $\vec{y}$,
- $\vec{y}$ cannot be expressed as a linear combination of $\vec{x}$

are both satisfied.

(c) give an example of $\vec{x}, \vec{y}$ such that none of them can be expressed as a linear combination of the other.

[**Note**: One vector, $\vec{u}$, being a linear combination of another vector, $\vec{v}$, simply means that $\vec{u}$ is a scalar multiple of $\vec{v}$.]

1.2.7 Consider three distinct nonzero vectors $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$. If possible,

(a) give an example of $\vec{x}, \vec{y}, \vec{z}$ such that each of them can be expressed as a linear combination of the other two.

(b) give an example of $\vec{x}, \vec{y}, \vec{z}$ such that the conditions

- $\vec{x}$ can be expressed as a linear combination of $\vec{y}$ and $\vec{z}$,
- $\vec{y}$ can be expressed as a linear combination of $\vec{x}$ and $\vec{z}$,
- $\vec{z}$ cannot be expressed as a linear combination of $\vec{x}$ and $\vec{y}$

are all satisfied.

(c) give an example of $\vec{x}, \vec{y}, \vec{z}$ such that the conditions

- $\vec{x}$ can be expressed as a linear combination of $\vec{y}$ and $\vec{z}$,
- $\vec{y}$ cannot be expressed as a linear combination of $\vec{x}$ and $\vec{z}$,
- $\vec{z}$ cannot be expressed as a linear combination of $\vec{x}$ and $\vec{y}$

are all satisfied.

(d) give an example of $\vec{x}, \vec{y}, \vec{z}$ such that none of them can be expressed as a linear combination of the other two.

1.3 The Norm and the Dot Product

Having introduced vectors in $\mathbb{R}^n$, the algebraic operations of addition and scalar multiplication along with their geometric interpretations, we now define the norm of a vector.

Definition 1.3.1
Norm

The **norm** (also known as **length** or **magnitude**) of $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ is the nonnegative real number

$$\|\vec{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Figure 1.3.1 shows that the norm of a vector in $\mathbb{R}^2$ represents the length or magnitude of the vector. This interpretation also applies to vectors in $\mathbb{R}^n$.

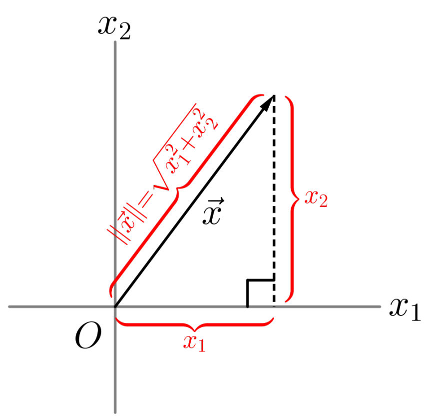


Figure 1.3.1: A vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$ and its norm, interpreted as length.

Example 1.3.2

For $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$, we have

$$\|\vec{x}\| = \sqrt{1^2 + 2^2} = \sqrt{5},$$

and for $\vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \in \mathbb{R}^4$, we have

$$\|\vec{y}\| = \sqrt{1^2 + 1^2 + 1^2 + 1^2} = \sqrt{4} = 2.$$

Exercise 4

Give examples of two different vectors in $\mathbb{R}^2$ whose norms are 1.

In Figure 1.3.1, the vector $\vec{x}$ was in standard position. Thus we may interpret $\|\vec{x}\|$ as the distance from the origin to the "tip" of $\vec{x}$. If our vector is instead from a point A to a point B, then we can think of $\|\vec{x}\|$ as the distance between the points A and B as illustrated in Figure 1.3.2.

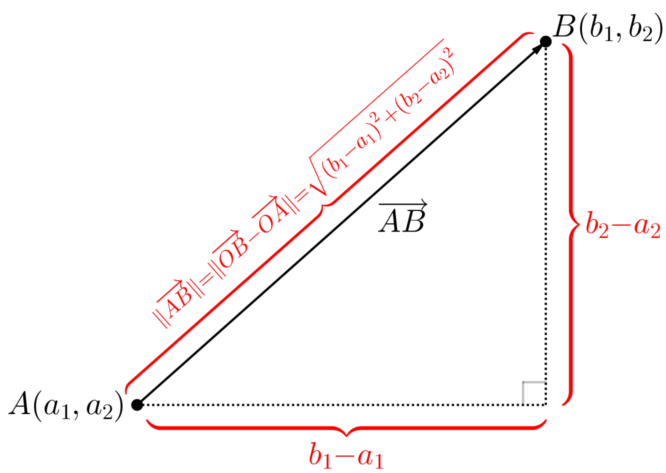


Figure 1.3.2: Viewing the norm between two points A and B in $\mathbb{R}^2$ as the distance between them. The picture in $\mathbb{R}^n$ is similar.

Example 1.3.3

Find the distance from $A(1, -1, 2)$ to $B(3, 2, 1)$.

Solution: Since

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix},$$

the distance from A to B is

$$\|\overrightarrow{AB}\| = \sqrt{2^2 + 3^2 + (-1)^2} = \sqrt{4 + 9 + 1} = \sqrt{14}.$$

The next theorem states some useful properties the norm obeys. We will employ these properties when we derive new results that rely on norms.

Theorem 1.3.4

(Properties of the Norm)

Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then

(a) $\|\vec{x}\| \geq 0$ with equality if and only if $\vec{x} = \vec{0}$.

(b) $\|c\vec{x}\| = |c|\|\vec{x}\|$.

(c) $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (the *Triangle Inequality*).

Property (c) is known as the Triangle Inequality and has a very nice geometric interpretation. Namely, that in the triangle determined by vectors $\vec{x}$, $\vec{y}$ and $\vec{x} + \vec{y}$ (see Figure

1.3.3), the length of any one side of the triangle cannot exceed the sum of the lengths of the remaining two sides. Or, more colloquially, *the shortest distance between two points is a straight line.*

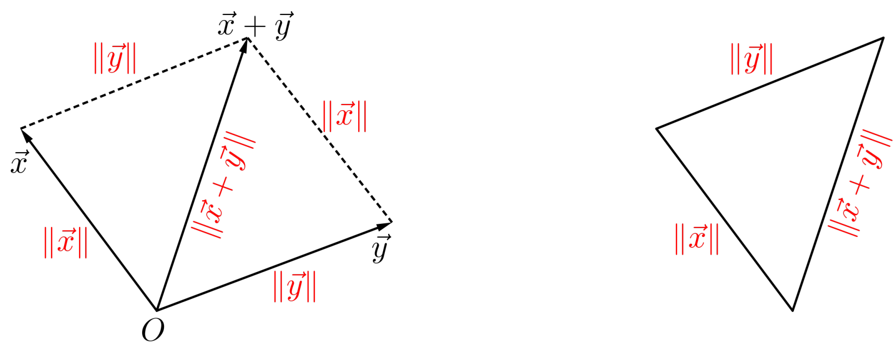


Figure 1.3.3: Interpreting the Triangle Inequality.

Definition 1.3.5
Unit Vector

A vector $\vec{x} \in \mathbb{R}^n$ is a **unit vector** if $\|\vec{x}\| = 1$.

Example 1.3.6

For instance,

- $\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in \mathbb{R}^2$ is a unit vector since $\|\vec{x}\| = \sqrt{1^2 + 0^2} = 1$
- $\vec{x} = -\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \mathbb{R}^3$ is a unit vector since $\|\vec{x}\| = \left| -\frac{1}{\sqrt{3}} \right| \sqrt{1^2 + 1^2 + 1^2} = \frac{1}{\sqrt{3}}\sqrt{3} = 1$
- $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2$ is **not** a unit vector since $\|\vec{x}\| = \sqrt{1^2 + 1^2} = \sqrt{2} \neq 1$.

Exercise 5

Let $\vec{x} \in \mathbb{R}^n$ be a unit vector and let $c \in \mathbb{R}$. Prove that if $c\vec{x}$ is a unit vector then $c = \pm 1$.

We will now show that, given a non-zero vector $\vec{x} \in \mathbb{R}^n$, there is a unit vector parallel to $\vec{x}$. First, we must define what we mean by parallel.

Definition 1.3.7
Parallel Vectors

Two nonzero vectors in $\mathbb{R}^n$ are **parallel** if they are scalar multiples of one another.

Example 1.3.8

The vectors

$$\vec{x} = \begin{bmatrix} 2 \\ -5 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} -4 \\ 10 \end{bmatrix}$$

are parallel since $\vec{y} = -2\vec{x}$, or equivalently, $\vec{x} = -\frac{1}{2}\vec{y}$. The vectors

$$\vec{u} = \begin{bmatrix} -2 \\ -3 \\ -4 \end{bmatrix} \quad \text{and} \quad \vec{v} = \begin{bmatrix} -2 \\ -1 \\ -13 \end{bmatrix}$$

are not parallel for $\vec{u} = c\vec{v}$ would imply that $-2 = -2c$, $-3 = -c$ and $-4 = -13c$ which implies that $c = 1, 3, \frac{4}{13}$ simultaneously, which is impossible.

Now, given a nonzero vector $\vec{x} \in \mathbb{R}^n$, the vector

$$\vec{y} = \frac{1}{\|\vec{x}\|}\vec{x}$$

is a unit vector parallel to $\vec{x}$. To see this, note that since $\vec{x} \neq \vec{0}$, we have $\|\vec{x}\| > 0$ by Theorem 1.3.4(a) and it follows that $1/\|\vec{x}\| > 0$. Thus $\vec{y}$ is a positive scalar multiple of $\vec{x}$. (Geometrically, we think of $\vec{y}$ as "pointing in the same direction" as $\vec{x}$.) Now

$$\|\vec{y}\,\| = \left\| \frac{1}{\|\vec{x}\|}\,\vec{x} \right\| = \left| \frac{1}{\|\vec{x}\|} \right| \|\vec{x}\| = \frac{1}{\|\vec{x}\|}\,\|\vec{x}\| = 1$$

so $\vec{y}$ is a unit vector parallel to $\vec{x}$. This derivation motivates the following definition.

Definition 1.3.9
Normalization

Given a nonzero vector $\vec{x} \in \mathbb{R}^n$, the vector

$$\widehat{x} = \frac{1}{\|\vec{x}\|}\vec{x}$$

is called the **normalization** of $\vec{x}$.

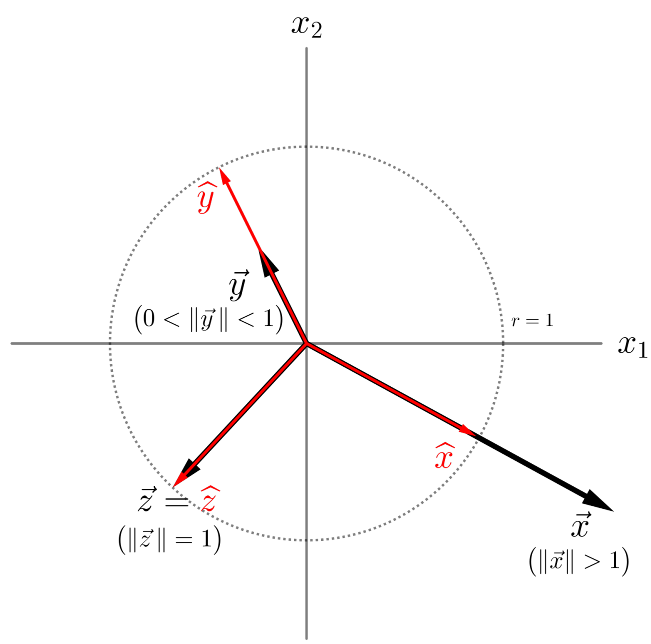


Figure 1.3.4: Three vectors $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^2$ and their normalizations $\widehat{x}$, $\widehat{y}$, and $\widehat{z}$.

Note that there are *two* unit vectors that are parallel to any given nonzero vector $\vec{x} \in \mathbb{R}^n$. Namely, the normalization $\widehat{x}$ of $\vec{x}$ and its negative, $-\widehat{x}$.

Example 1.3.10

Find a unit vector parallel to $\vec{x} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$.

Solution: What we we want here is the normalization of $\vec{x}$.

Since $\|\vec{x}\| = \sqrt{4^2+5^2+6^2} = \sqrt{16+25+36} = \sqrt{77}$, we have

$$\widehat{x} = \frac{1}{\sqrt{77}} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4/\sqrt{77} \\ 5/\sqrt{77} \\ 6/\sqrt{77} \end{bmatrix}$$

is the desired vector.

(Of course, $-\widehat{x}$ is also an acceptable answer.)

We now define the dot product of two vectors in $\mathbb{R}^n$. We will see how this product is related to the norm, and use it to compute the angles between nonzero vectors.

Definition 1.3.11
Dot Product

Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$. The **dot product** of $\vec{x}$ and $\vec{y}$ is the real number

$$\vec{x} \cdot \vec{y} = x_1 y_1 + \cdots + x_n y_n.$$

The dot product is sometimes referred to the *scalar product* or the *standard inner product*. The term scalar product comes from the fact that give two vectors in $\mathbb{R}^n$, their dot product returns a real number, which we call a *scalar*.

Example 1.3.12

We have

- $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ -4 \\ 5 \end{bmatrix} = 1(-3) + 1(-4) + 2(5) = -3 - 4 + 10 = 3.$
- $\begin{bmatrix} 1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \end{bmatrix} = 1(-2) + 2(1) = -2 + 2 = 0.$

Notice that the dot product of two non-zero vectors can be zero.

Exercise 6

In $\mathbb{R}^4$, let

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \vec{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

that is, for $i = 1, \ldots, 4$, $\vec{e}_i$ is the vector with an entry of 1 in the ith component and zeros elsewhere.

Determine $\vec{e}_i \cdot \vec{e}_j$. [**Hint:** Your answer should depend on i and j.]

The next theorem states some useful properties of the dot product.

Theorem 1.3.13 **(Properties of the Dot Product)**

Let $\vec{w}, \vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

(a) $\vec{x} \cdot \vec{y} \in \mathbb{R}$.

(b) $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$.

(c) $\vec{x} \cdot \vec{0} = 0$.

(d) $\vec{x} \cdot \vec{x} = \|\vec{x}\|^2$.

(e) $(c\vec{x}) \cdot \vec{y} = c(\vec{x} \cdot \vec{y}) = \vec{x} \cdot (c\vec{y})$.

(f) $\vec{w} \cdot (\vec{x} \pm \vec{y}) = \vec{w} \cdot \vec{x} \pm \vec{w} \cdot \vec{y}$.

Proof: We prove (b), (d) and (e). Let $c \in \mathbb{R}$ and let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$. For (b) we have

$$\vec{x} \cdot \vec{y} = x_1y_1 + \cdots + x_ny_n = y_1x_1 + \cdots + y_nx_n = \vec{y} \cdot \vec{x}.$$

Now to prove (d), we have

$$\vec{x} \cdot \vec{x} = x_1x_1 + \cdots + x_nx_n = x_1^2 + \cdots + x_n^2 = \|\vec{x}\|^2.$$

For (e),

$$(c\vec{x}) \cdot \vec{y} = (cx_1)y_1 + \cdots + (cx_n)y_n = c(x_1y_1 + \cdots + x_ny_n) = c(\vec{x} \cdot \vec{y}).$$

That $\vec{x} \cdot (c\vec{y}) = c(\vec{x} \cdot \vec{y})$ is shown similarly. □

Theorem 1.3.13(d) is significant as it provides a connection between the dot product and the norm which can be useful when verifying statements involving norms. The following inequality, known as the Cauchy–Schwarz Inequality[4], gives another relationship between the dot product and the norm that we will make use of later in this section. Note that the proof of the Cauchy–Schwarz Inequality makes use of Theorem 1.3.13(d).

Theorem 1.3.14 **(Cauchy–Schwarz Inequality)**

For any two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$, we have

$$|\vec{x} \cdot \vec{y}\,| \leq \|\vec{x}\|\|\vec{y}\|.$$

[4] Also known as the Cauchy–Bunyakovsky–Schwarz Inequality.

Proof: Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Using Theorem 1.3.13, we have

$$\begin{aligned}
0 &\leq \|\alpha\vec{x} - \vec{y}\|^2 \\
&= (\alpha\vec{x} - \vec{y}) \cdot (\alpha\vec{x} - \vec{y}) && \text{by (d)} \\
&= (\alpha\vec{x}) \cdot (\alpha\vec{x}) + (\alpha\vec{x}) \cdot (-\vec{y}) + (-\vec{y}) \cdot (\alpha\vec{x}) + (-\vec{y}) \cdot (-\vec{y}) && \text{by (b), (f)} \\
&= \alpha^2(\vec{x} \cdot \vec{x}) - \alpha(\vec{x} \cdot \vec{y}) - \alpha(\vec{y} \cdot \vec{x}) + \vec{y} \cdot \vec{y} && \text{by (e)} \\
&= \alpha^2\|\vec{x}\|^2 - 2\alpha(\vec{x} \cdot \vec{y}) + \|\vec{y}\|^2 && \text{by (b), (d).}
\end{aligned}$$

We see that

$$0 \leq \|\vec{x}\|^2\alpha^2 - 2(\vec{x} \cdot \vec{y})\alpha + \|\vec{y}\|^2. \tag{1.1}$$

Note that $\|\vec{x}\|^2\alpha^2 - 2(\vec{x} \cdot \vec{y})\alpha + \|\vec{y}\|^2$ is a quadratic polynomial in the variable α. Equation (1.1) shows that this polynomial is nonnegative and thus it has at most one real root. It follows that its discriminant[5] cannot be positive. Thus

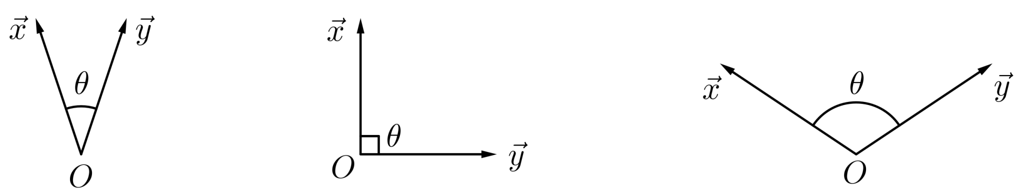
$$0 \geq \big(-2(\vec{x} \cdot \vec{y})\big)^2 - 4\|\vec{x}\|^2\|\vec{y}\|^2 = 4(\vec{x} \cdot \vec{y})^2 - 4\|\vec{x}\|^2\|\vec{y}\|^2.$$

Rearranging and dividing by 4 gives $(\vec{x} \cdot \vec{y})^2 \leq \|\vec{x}\|^2\|\vec{y}\|^2$. Finally, taking square roots gives $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$ as required. □

Exercise 7

Let $\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and let $\vec{y} \in \mathbb{R}^3$ be such that $\vec{x} \cdot \vec{y} = -7$. What are the possible values of $\|\vec{y}\|$?

We now look at how norms and dot products lead to a nice geometric interpretation about angles between vectors. Given two nonzero vectors $\vec{x}, \vec{y} \in \mathbb{R}^2$, they determine an angle θ as shown in Figure 1.3.5. We restrict θ to $0 \leq \theta \leq \pi$ to avoid multiple values for θ and to avoid reflex angles.



(a) Acute: $0 \leq \theta < \frac{\pi}{2}$ (b) Perpendicular: $\theta = \frac{\pi}{2}$ (c) Obtuse: $\frac{\pi}{2} < \theta \leq \pi$

Figure 1.3.5: Every two nonzero vectors in $\mathbb{R}^2$ either determine an acute angle, are perpendicular, or determine an obtuse angle.

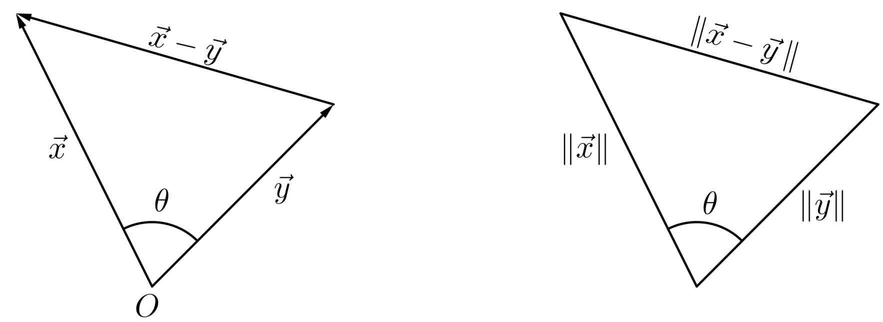
Theorem 1.3.15

For two nonzero vectors $\vec{x}, \vec{y} \in \mathbb{R}^2$ determining an angle θ,

$$\vec{x} \cdot \vec{y} = \|\vec{x}\|\|\vec{y}\| \cos\theta.$$

[5] Recall that for a quadratic polynomial $ax^2 + bx + c$, the discriminant is $b^2 - 4ac$.

Proof: Consider the triangle determined by the vectors $\vec{x}$, $\vec{y}$ and $\vec{x}-\vec{y}$.



From the Cosine Law, we have

$$\|\vec{x}-\vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\|\vec{x}\|\|\vec{y}\|\cos\theta. \tag{1.2}$$

Using Theorem 1.3.13, we obtain

$$\begin{aligned}\|\vec{x}-\vec{y}\|^2 &= (\vec{x}-\vec{y})\cdot(\vec{x}-\vec{y}) && \text{by (d)}\\ &= \vec{x}\cdot\vec{x} - \vec{y}\cdot\vec{x} - \vec{x}\cdot\vec{y} + \vec{y}\cdot\vec{y} && \text{by (e), (f)}\\ &= \|\vec{x}\|^2 - 2(\vec{x}\cdot\vec{y}) + \|\vec{y}\|^2 && \text{by (b), (d).}\end{aligned}$$

Thus (1.2) becomes

$$\|\vec{x}\|^2 - 2(\vec{x}\cdot\vec{y}) + \|\vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2\|\vec{x}\|\|\vec{y}\|\cos\theta$$

and subtracting $\|\vec{x}\|^2 + \|\vec{y}\|^2$ from both sides and then multiplying both sides by $-\frac{1}{2}$ gives $\vec{x}\cdot\vec{y} = \|\vec{x}\|\|\vec{y}\|\cos\theta$ as required. □

Theorem 1.3.15 gives a relationship between the angle θ determined by two nonzero vectors $\vec{x},\vec{y}\in\mathbb{R}^2$ and their dot product. This relationship motivates us to define the angle determined by two vectors in $\mathbb{R}^n$.

Definition 1.3.16

Angle Determined by Two Vectors in $\mathbb{R}^n$

Let $\vec{x},\vec{y}\in\mathbb{R}^n$ be two nonzero vectors. The **angle** θ they determine (with $0\le\theta\le\pi$) is such that

$$\vec{x}\cdot\vec{y} = \|\vec{x}\|\|\vec{y}\|\cos\theta.$$

The equation in Definition 1.3.16 can be rearranged as

$$\cos\theta = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\|\vec{y}\|} \tag{1.3}$$

which will allow us to explicitly solve for θ (again, we are assuming $\vec{x},\vec{y}\neq\vec{0}$).

Note that we have simply *defined* the angle between two nonzero vectors in $\mathbb{R}^n$ whereas we *derived* the angle in $\mathbb{R}^2$. It is natural to ask if Definition 1.3.16 is reasonable. Recall that for $\vec{x}, \vec{y} \in \mathbb{R}^n$, the Cauchy–Schwarz Inequality states that $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$. Rearranging gives

$$-\|\vec{x}\|\|\vec{y}\| \leq \vec{x} \cdot \vec{y} \leq \|\vec{x}\|\|\vec{y}\|,$$

and for $\vec{x}, \vec{y} \neq \vec{0}$, we obtain

$$-1 \leq \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|} \leq 1.$$

Thus, it is reasonable that the quantity $\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|}$ be equated to $\cos\theta$ since $-1 \leq \cos\theta \leq 1$. Thus (1.3), and ultimately Definition 1.3.16, make sense.

We can now use equation (1.3) solve for θ:

$$\theta = \arccos\left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|}\right).$$

Example 1.3.17 Compute the angle determined by the vectors $\vec{x} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$.

Solution: We have that

$$\cos\theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \frac{2(1) + 1(-1) - 1(-2)}{\sqrt{4+1+1}\sqrt{1+1+4}} = \frac{3}{\sqrt{6}\sqrt{6}} = \frac{1}{2}$$

so

$$\theta = \arccos\left(\frac{1}{2}\right) = \frac{\pi}{3}.$$

Exercise 8 Compute the angle determined by the vectors $\vec{x} = \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 1 \end{smallmatrix}\right]$ and $\vec{y} = \left[\begin{smallmatrix} 2 \\ 0 \\ 0 \\ 2 \end{smallmatrix}\right]$.

For nonzero vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ determining an angle θ, we are often not interested in the specific value of θ, but rather in the approximate size of θ. That is, we are often only concerned if $\vec{x}$ and $\vec{y}$ determine an acute angle, an obtuse angle, or if the vectors are perpendicular (refer back to Figure 1.3.5).

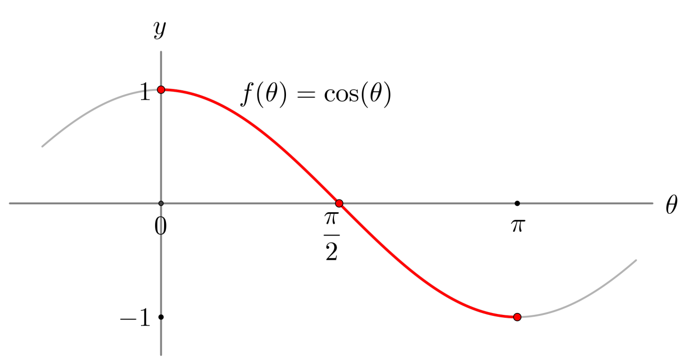


Figure 1.3.6: The graph of $f(\theta) = \cos(\theta)$ for $0 \leq \theta \leq \pi$.

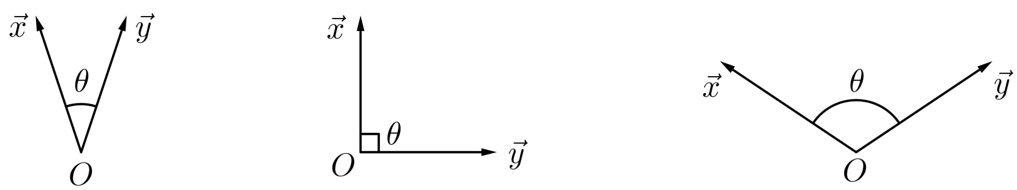
Using the graph of $f(\theta) = \cos(\theta)$ given in Figure 1.3.6, we see that

$$\cos\theta > 0 \text{ for } 0 \leq \theta < \frac{\pi}{2},$$
$$\cos\theta = 0 \text{ for } \theta = \frac{\pi}{2},$$
$$\cos\theta < 0 \text{ for } \frac{\pi}{2} < \theta \leq \pi.$$

It then follows from (1.3) that the sign of $\cos\theta$ is determined by the sign of $\vec{x} \cdot \vec{y}$ since $\|\vec{x}\|\|\vec{y}\| > 0$. Thus

$$\begin{array}{lcccl} \vec{x} \cdot \vec{y} > 0 & \Longleftrightarrow & 0 \leq \theta < \frac{\pi}{2} & \Longleftrightarrow & \vec{x} \text{ and } \vec{y} \text{ determine an acute angle,} \\ \vec{x} \cdot \vec{y} = 0 & \Longleftrightarrow & \theta = \frac{\pi}{2} & \Longleftrightarrow & \vec{x} \text{ and } \vec{y} \text{ are perpendicular,} \\ \vec{x} \cdot \vec{y} < 0 & \Longleftrightarrow & \frac{\pi}{2} < \theta \leq \pi & \Longleftrightarrow & \vec{x} \text{ and } \vec{y} \text{ determine an obtuse angle.} \end{array}$$

This is illustrated in Figure 1.3.7.



(a) Acute: $\vec{x} \cdot \vec{y} > 0$ (b) Perpendicular: $\vec{x} \cdot \vec{y} = 0$ (c) Obtuse: $\vec{x} \cdot \vec{y} < 0$

Figure 1.3.7: The dot product of two nonzero vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ tells us if they determine an acute angle, are perpendicular, or if they determine an obtuse angle.

Example 1.3.18 For

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 6 \\ -2 \end{bmatrix},$$

we compute

$$\vec{x} \cdot \vec{y} = 1(6) + 2(-2) = 2 > 0$$

and so $\vec{x}$ and $\vec{y}$ determine an acute angle.

Note that to find the exact angle determined by $\vec{x}$ and $\vec{y}$ in the previous example we compute

$$\cos\theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \frac{2}{\sqrt{1+4}\sqrt{36+4}} = \frac{2}{\sqrt{5}\sqrt{40}} = \frac{2}{\sqrt{200}} = \frac{2}{10\sqrt{2}} = \frac{1}{5\sqrt{2}}$$

so

$$\theta = \arccos\left(\frac{1}{5\sqrt{2}}\right)$$

which is our exact answer for θ. As a decimal number rounded to the nearest millionth, we have $\theta \approx 1.428899$, but that this is an approximation rather than the exact value. In this course, it is normally expected that students give exact answers unless otherwise stated.

We have defined the norm for any vector in $\mathbb{R}^n$ and the dot product for any two vectors in $\mathbb{R}^n$. Our resulting work with angles determined by two vectors has required that our vectors be nonzero, but we do not wish to continue excluding the zero vector. Since $\vec{x} \cdot \vec{0} = 0$ for every $\vec{x} \in \mathbb{R}^n$, it would seem natural to say that the zero vector is perpendicular to every vector. However, the word perpendicular is a geometric term meaning to make a right angle, and the zero vector does not make any angle with any vector. We thus make the following definition.

Definition 1.3.19
Orthogonal

Two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ are said to be **orthogonal** if $\vec{x} \cdot \vec{y} = 0$.

Thus if $\vec{x}, \vec{y} \in \mathbb{R}^n$ are nonzero, then they are orthogonal exactly when they are perpendicular. However, if either of $\vec{x}, \vec{y}$ are the zero vector, then we will say they are orthogonal, but we cannot say they are perpendicular since

$$\cos\theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|}$$

is not defined if either $\vec{x}$ or $\vec{y}$ is the zero vector. Thus we interpret $\vec{x}$ and $\vec{y}$ being orthogonal to mean that their dot product is zero, and if both $\vec{x}$ and $\vec{y}$ are nonzero, then they are perpendicular and determine an angle of $\frac{\pi}{2}$.

Example 1.3.20

For instance,

- $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ are orthogonal since $\vec{x} \cdot \vec{y} = 1(2) + 2(-1) = 0$.
- $\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ are **not** orthogonal since $\vec{x} \cdot \vec{y} = (1)(1)+(1)(2)+(1)(3) = 6 \neq 0$.

Exercise 9

Give an example of a nonzero vector $\vec{x} \in \mathbb{R}^3$ that is orthogonal to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Section 1.3 Problems

1.3.1. Let
$$\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}, \quad \vec{z} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{w} = \begin{bmatrix} 2 \\ -2 \\ -2 \end{bmatrix}.$$

(a) Compute the norm of each of the vectors.

(b) Compute the dot product of each pair of distinct vectors.

(c) For each pair of distinct vectors, determine if they form an acute angle, an obtuse angle, or are orthogonal.

(d) Find the exact radian angle θ, $0 \leq \theta \leq \pi$ determined by $\vec{x}$ and $\vec{y}$.

(e) Find a unit vector $\vec{u}$ in the opposite direction of $\vec{x} + \vec{w}$.

1.3.2. Let $k \in \mathbb{R}$ and let
$$\vec{u} = \begin{bmatrix} 5 \\ k \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{v} = \begin{bmatrix} -k \\ k \\ 2 \end{bmatrix}$$

Find the following if possible.

(a) All k such that $\vec{u}$ and $\vec{v}$ are orthogonal.

(b) All k such that $\vec{u}$ and $\vec{v}$ determine an acute angle.

(c) All k such that $\vec{u}$ and $\vec{v}$ determine an obtuse angle.

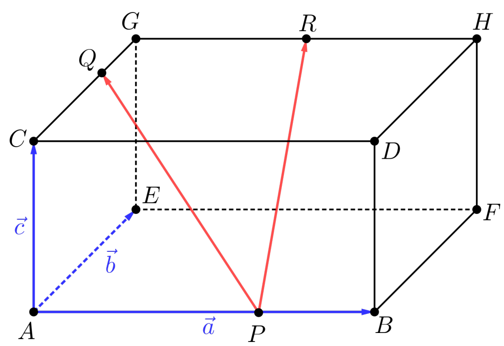
(d) All k such that $\vec{u}$ and $\vec{v}$ are parallel.

1.3.3. Consider the quadrilateral $PQRS$ with vertices $P(1,2,3)$, $Q(2,1,5)$, $R(4,1,4)$, and $S(3,2,2)$ (the "name" $PQRS$ implies that edges of the quadrilateral are the segments PQ, QR, RS and SP).

(a) Show that $PQRS$ is a rectangle.

(b) Find the area of this rectangle.

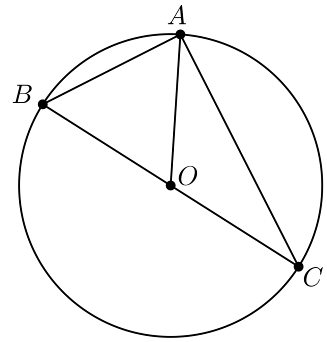
1.3.4. Consider the rectangular box with vertices A, B, C, D, E, F, G, H as shown below. Let $\vec{a} = \overrightarrow{AB}$, $\vec{b} = \overrightarrow{AE}$ and $\vec{c} = \overrightarrow{AC}$ with $\|\vec{a}\| = 6$, $\|\vec{b}\| = 3$, $\|\vec{c}\| = 4$ and any two of $\vec{a}, \vec{b}, \vec{c}$ being orthogonal. The distance between A and P is 4, the distance between C and Q is 2 and R is the midpoint of G and H. You may not make any assumptions on the values of $\vec{a}$, $\vec{b}$ and $\vec{c}$.



(a) Find $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ in terms of $\vec{a}$, $\vec{b}$ and $\vec{c}$. Simplify your answers.

(b) Find the cosine of the angle determined by $\overrightarrow{PQ}$ and $\overrightarrow{PR}$.

1.3.5. Consider a circle centred at a point O. Let B and C be two points on this circle such that O lies on the line segment connecting B and C. Let A be any point on the circle. Using vectors, show that $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are orthogonal.



1.3.6. Let $\vec{u}, \vec{v} \in \mathbb{R}^n$ be such that $\vec{v} = k\vec{u}$ for some $k \in \mathbb{R}$ with $k \geq 0$. Prove that $\|\vec{u} + \vec{v}\| = \|\vec{u}\| + \|\vec{v}\|$.

1.3.7. Let $\vec{u}, \vec{v} \in \mathbb{R}^n$. Prove:

(a) If $\vec{u} - \vec{v}$ and $\vec{u} + \vec{v}$ are orthogonal then $\|\vec{u}\| = \|\vec{v}\|$.

(b) If $\|\vec{u}\| = \|\vec{v}\|$ then $\vec{u} - \vec{v}$ and $\vec{u} + \vec{v}$ are orthogonal.

1.3.8. Recall the Cauchy–Schwarz Inequality, which states that for any $\vec{x}, \vec{y} \in \mathbb{R}^n$, $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$.

(a) Show that if either of $\vec{x}$ or $\vec{y}$ are the zero vector, then $|\vec{x} \cdot \vec{y}| = \|\vec{x}\|\|\vec{y}\|$.

(b) Show that for nonzero $\vec{x}, \vec{y} \in \mathbb{R}^n$, $|\vec{x} \cdot \vec{y}| = \|\vec{x}\|\|\vec{y}\|$ if and only if $\vec{x}$ and $\vec{y}$ are parallel. Hint: consider the possible angles $\vec{x}$ and $\vec{y}$ can determine if they are parallel.

(c) Show that

$$|x_1y_1 + x_2y_2 + x_3y_3| \leq \sqrt{x_1^2 + x_2^2 + x_3^2}\sqrt{y_1^2 + y_2^2 + y_3^2}.$$

for all $x_1, x_2, x_3, y_1, y_2, y_3 \in \mathbb{R}$.

1.3.9. Prove the Triangle Inequality, that is, show that for $\vec{x}, \vec{y} \in \mathbb{R}^n$,

$$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|.$$

[**Hint:** First show that $\|\vec{x}+\vec{y}\|^2 \leq \left(\|\vec{x}\|+\|\vec{y}\|\right)^2$ by using the Cauchy–Schwarz Inequality.]

1.3.10. Define a new norm for vectors $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ as follows:

$$\|\vec{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|.$$

(a) For each of

(i) $\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (ii) $\vec{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ (iii) $\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 6 \end{bmatrix}$ (iv) $\vec{x} = \begin{bmatrix} 1 \\ -3 \\ 2 \\ -1 \end{bmatrix}$

evaluate both $\|\vec{x}\|$ and $\|\vec{x}\|_1$.

(b) Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Show that our above definition satisfies the following properties:

(i) $\|\vec{x}\|_1 \geq 0$ with equality if and only if $\vec{x} = \vec{0}$.
(ii) $\|c\vec{x}\|_1 = |c|\|\vec{x}\|_1$.
(iii) $\|\vec{x} + \vec{y}\|_1 \leq \|\vec{x}\|_1 + \|\vec{y}\|_1$.

(c) For any nonzero $\vec{x} \in \mathbb{R}^n$, let $\vec{y} = \dfrac{1}{\|\vec{x}\|_1}\vec{x}$. Show that $\|\vec{y}\|_1 = 1$.

(d) Find a vector $\vec{y} \in \mathbb{R}^3$ in same the direction of $\vec{x} = \left[\begin{smallmatrix} 1 \\ -2 \\ 6 \end{smallmatrix}\right]$ such that $\|\vec{y}\|_1 = 1$.

(e) In $\mathbb{R}^2$, plot all points P such that $\|\overrightarrow{OP}\|_1 = 1$.
Hint: Do this separately in each quadrant and remember that for any $a \in \mathbb{R}$,

$$|a| = \begin{cases} a, & a \geq 0 \\ -a, & a < 0 \end{cases}.$$

1.3.11. Define a new norm for vectors $\vec{x} = \left[\begin{smallmatrix} x_1 \\ \vdots \\ x_n \end{smallmatrix}\right] \in \mathbb{R}^n$ as follows:

$$\|\vec{x}\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\} = \max_{1 \leq i \leq n} |x_i|.$$

(a) For each of

(i) $\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (ii) $\vec{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ (iii) $\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 6 \end{bmatrix}$ (iv) $\vec{x} = \begin{bmatrix} 1 \\ -3 \\ 2 \\ -1 \end{bmatrix}$

evaluate both $\|\vec{x}\|$ and $\|\vec{x}\|_\infty$.

(b) Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Show that our above definition satisfies the following properties:

(i) $\|\vec{x}\|_\infty \geq 0$ with equality if and only if $\vec{x} = \vec{0}$.
(ii) $\|c\vec{x}\|_\infty = |c|\|\vec{x}\|_\infty$.
(iii) $\|\vec{x} + \vec{y}\|_\infty \leq \|\vec{x}\|_\infty + \|\vec{y}\|_\infty$.

(c) For any nonzero $\vec{x} \in \mathbb{R}^n$, let $\vec{y} = \dfrac{1}{\|\vec{x}\|_\infty}\vec{x}$. Show that $\|\vec{y}\|_\infty = 1$.

(d) Find a vector $\vec{y} \in \mathbb{R}^3$ in same the direction of $\vec{x} = \left[\begin{smallmatrix} 1 \\ -2 \\ 6 \end{smallmatrix}\right]$ such that $\|\vec{y}\|_\infty = 1$.

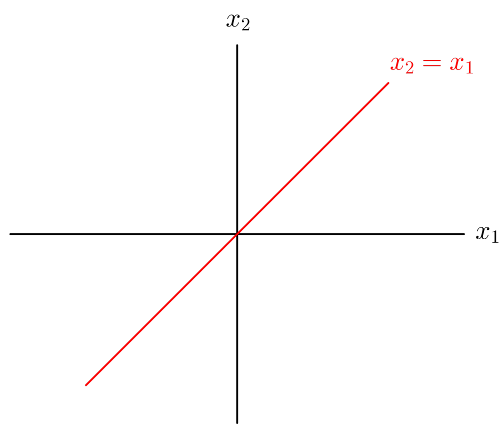
(e) In $\mathbb{R}^2$, plot all those points P such that $\|\overrightarrow{OP}\|_\infty = 1$.

1.4 Vector Equations of Lines and Planes

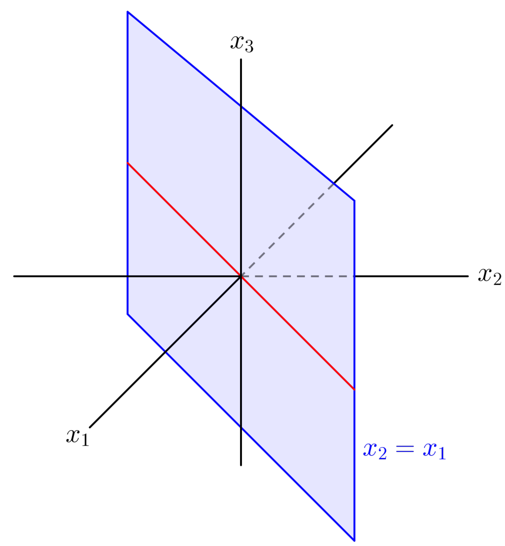
In $\mathbb{R}^2$, we define lines by equations such as

$$x_2 = mx_1 + b \quad \text{or} \quad ax_1 + bx_2 = c$$

where m, a, b, c are constants. How do we describe lines in $\mathbb{R}^n$ (for example, in $\mathbb{R}^3$)? It might be tempting to think the above equations are also equations of lines in $\mathbb{R}^n$ as well, but this is not the case. Consider the graph of the equation $x_2 = x_1$ in $\mathbb{R}^2$. This graph consists of all points (x_1, x_2) such that $x_2 = x_1$, which yields a line (see Figure 1.4.1a). If we consider the equation $x_2 = x_1$ in $\mathbb{R}^3$, then we are considering all points (x_1, x_2, x_3) with the property that $x_2 = x_1$. Notice that there is no restriction on x_3, so we can take x_3 to be any real number. It follows that the equation $x_2 = x_1$ represents a vertical plane in $\mathbb{R}^3$ and not a line (see Figure 1.4.1b).



(a) The graph of $x_2 = x_1$ is a line in $\mathbb{R}^2$.



(b) The graph of $x_2 = x_1$ is a plane in $\mathbb{R}^3$. The red line indicates the intersection of this plane with the x_1x_2-plane.

Figure 1.4.1: The equation $x_2 = x_1$ represents a line in $\mathbb{R}^2$, but it represents a plane in $\mathbb{R}^3$.

Note that to uniquely determine a line in $\mathbb{R}^n$, we must know two things:

- A point P on the line,
- A nonzero vector $\vec{d}$ in the direction of the line (called a direction vector for the line).

Definition 1.4.1

Vector Equation of a Line, Direction Vector

A line in $\mathbb{R}^n$ through a point P with direction $\vec{d}$, where $\vec{d} \in \mathbb{R}^n$ is nonzero, is given by the **vector equation**

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \overrightarrow{OP} + t\vec{d}, \quad t \in \mathbb{R}.$$

The vector $\vec{d}$ is called a **direction vector** for this line.

We can see from Figure 1.4.2 how the line through P with direction $\vec{d}$ is "drawn out" by the vector $\vec{x} = \overrightarrow{OP} + t\,\vec{d}$ as $t \in \mathbb{R}$ varies from $-\infty$ to ∞.

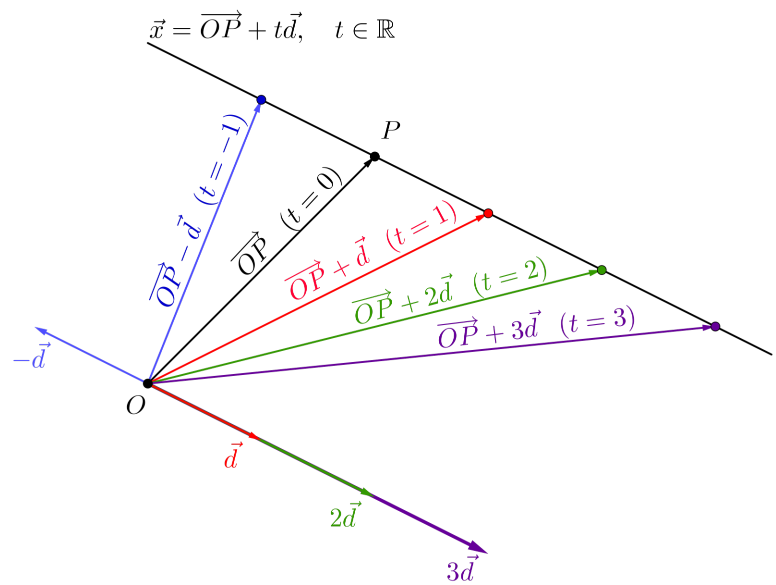


Figure 1.4.2: The line through P with direction $\vec{d}$ and the vector $\overrightarrow{OP} + t\,\vec{d}$ with some additional points plotted for a few values of $t \in \mathbb{R}$.

We can also think of the equation $\vec{x} = \overrightarrow{OP} + t\,\vec{d}$ as first moving us from the origin to the point P, and then moving from P as far as we like in the direction given by $\vec{d}$. This is shown in Figure 1.4.3.

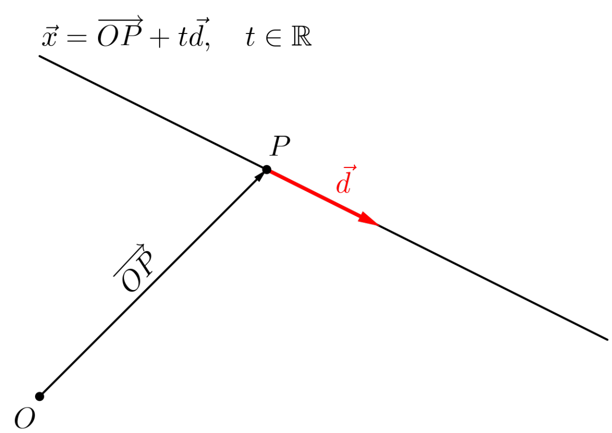


Figure 1.4.3: An equivalent way to understand the vector equation $\vec{x} = \overrightarrow{OP} + t\,\vec{d}$.

Example 1.4.2

Find a vector equation of the line through the points $A(1, 1, -1)$ and $B(4, 0, -3)$.

Solution: We first find a direction vector for the line. Since the line passes through the points A and B, we take the direction vector to be the vector from A to B. That is,

$$\vec{d} = \overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = \begin{bmatrix} 4 \\ 0 \\ -3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}.$$

Hence, using the point A, we have a vector equation for our line:

$$\vec{x} = \overrightarrow{OA} + t\overrightarrow{AB} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + t \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}, \quad t \in \mathbb{R}.$$

Note that a vector equation for a line is not unique. In fact, in Example 1.4.2, we could have used the vector $\overrightarrow{BA}$ as our direction vector, and we could have used B as the point on our line to obtain

$$\vec{x} = \overrightarrow{OB} + t\overrightarrow{BA} = \begin{bmatrix} 4 \\ 0 \\ -3 \end{bmatrix} + t \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \quad t \in \mathbb{R}.$$

Indeed, we can use any known point on the line and any nonzero scalar multiple of the direction vector for the line when constructing a vector equation. Thus, there are infinitely many vector equations for a line (see Figure 1.4.4).

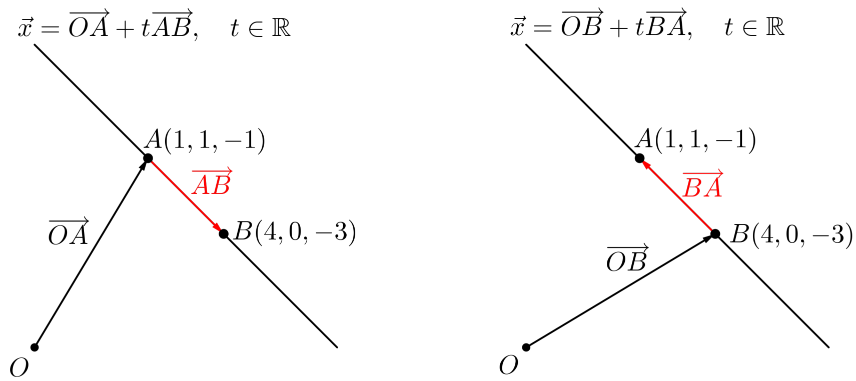


Figure 1.4.4: Two different vector equations for the same line.

Finally, given the vector equation we derived for the line in Example 1.4.2, we have

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + t \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 3t \\ -t \\ -2t \end{bmatrix} = \begin{bmatrix} 1+3t \\ 1-t \\ -1-2t \end{bmatrix}$$

from which it follows that

$$\begin{aligned} x_1 &= 1+3t \\ x_2 &= 1-t, \qquad t \in \mathbb{R} \\ x_3 &= -1-2t \end{aligned}$$

which we call *parametric equations* of the line. For each choice of $t \in \mathbb{R}$, these equations give the x_1-, x_2- and x_3-coordinates of a point on the line. Note that since a vector equation for a line is not unique, neither are the parametric equations for a line.

Exercise 10 Give both a vector equation and parametric equations for the line in $\mathbb{R}^3$ that passes through $P(2,-1,1)$ with direction vector $\vec{d} = \left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$.

We can easily extend the idea of a vector equation for a line in $\mathbb{R}^n$ to a vector equation for a plane in $\mathbb{R}^n$.

Definition 1.4.3

Vector Equation of a Plane

A **vector equation** for a plane in $\mathbb{R}^n$ through a point P is given by

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \overrightarrow{OP} + s\vec{u} + t\vec{v}, \quad s, t \in \mathbb{R}$$

where $\vec{u}, \vec{v} \in \mathbb{R}^n$ are nonzero nonparallel vectors.

We may think of this vector equation as taking us from the origin to the point P on the plane, and then adding any linear combination of $\vec{u}$ and $\vec{v}$ to reach any point on the plane. It is important to note that the parameters s and t are chosen independently of one another, that is, the choice of one parameter does not determine the choice of the other (see Figure 1.4.5).

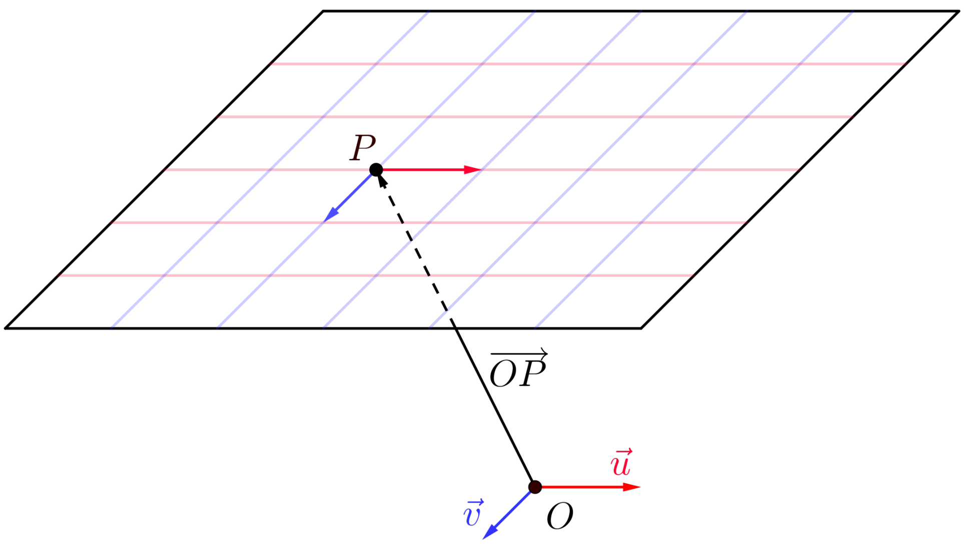


Figure 1.4.5: Using vectors to describe a plane in $\mathbb{R}^n$

Example 1.4.4

Find a vector equation for the plane containing the points $A(1, 1, 1)$, $B(1, 2, 3)$ and $C(-1, 1, 2)$.

Solution: We compute

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

$$\overrightarrow{AC} = \overrightarrow{OC} - \overrightarrow{OA} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

and note that $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are nonzero and nonparallel. A vector equation is thus

$$\vec{x} = \overrightarrow{OA} + s\overrightarrow{AB} + t\overrightarrow{AC} = \begin{bmatrix}1\\1\\1\end{bmatrix} + s\begin{bmatrix}0\\1\\2\end{bmatrix} + t\begin{bmatrix}-2\\0\\1\end{bmatrix}, \quad s,t \in \mathbb{R}.$$

Considering our vector equation from Example 1.4.4, we see that by setting either of $s,t \in \mathbb{R}$ to be zero and letting the other parameter be arbitrary, we obtain vector equations for two lines – each of which lie in the given plane:

$$\vec{x} = \overrightarrow{OA} + s\overrightarrow{AB} = \begin{bmatrix}1\\1\\1\end{bmatrix} + s\begin{bmatrix}0\\1\\2\end{bmatrix},\ s \in \mathbb{R} \quad \text{and} \quad \vec{x} = \overrightarrow{OA} + t\overrightarrow{AC} = \begin{bmatrix}1\\1\\1\end{bmatrix} + t\begin{bmatrix}-2\\0\\1\end{bmatrix},\ t \in \mathbb{R}.$$

This is illustrated in Figure 1.4.6.

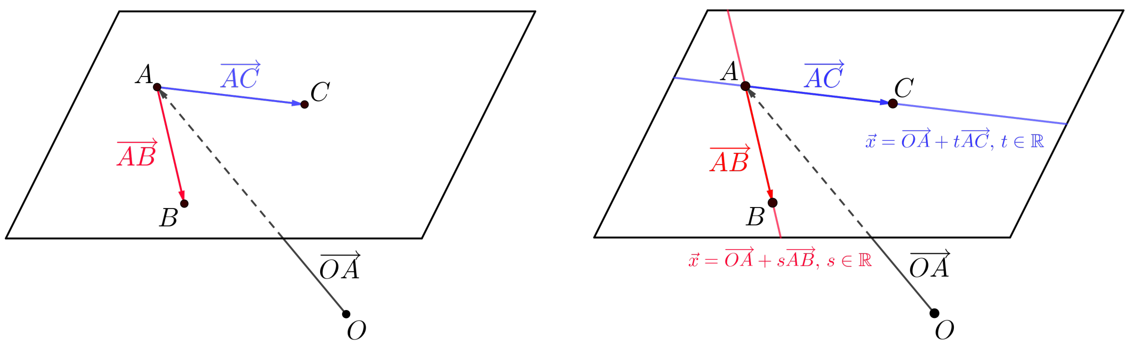


Figure 1.4.6: The plane through the points A, B and C with vector equation $\vec{x} = \overrightarrow{OA} + s\overrightarrow{AB} + t\overrightarrow{AC}$, $s,t \in \mathbb{R}$.

We also note that evaluating the right hand side of the vector equation derived in Example 1.4.4 gives

$$\vec{x} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}1\\1\\1\end{bmatrix} + s\begin{bmatrix}0\\1\\2\end{bmatrix} + t\begin{bmatrix}-2\\0\\1\end{bmatrix} = \begin{bmatrix}1-2t\\1+s\\1+2s+t\end{bmatrix}$$

from which we derive *parametric equations* of the plane:

$$\begin{aligned} x_1 &= 1-2t \\ x_2 &= 1+s \qquad\qquad s,t \in \mathbb{R}. \\ x_3 &= 1+2s+t \end{aligned}$$

It is worth observing that we require *two* parameters here, whereas we only required *one* parameter for the parametric equations of a line. This is tied to the fact that, geometrically, a plane is two-dimensional whereas a line is one-dimensional. For now, dimension is a concept that you should intuitively understand. We will give a more precise definition of dimension later in the course.

Finally, we note that as with lines, our vector equation for the plane in Example 1.4.4 is not unique as we could have chosen

$$\vec{x} = \overrightarrow{OB} + s\overrightarrow{BC} + t\overrightarrow{AB}, \quad s,t \in \mathbb{R}$$

as a vector equation instead (it is easy to verify that $\overrightarrow{BC}$ and $\overrightarrow{AB}$ are nonzero and nonparallel).

Example 1.4.5 Find a vector equation of the plane containing the point $P(1, -1, -2)$ and the line with vector equation

$$\vec{x} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} + r \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}, \quad r \in \mathbb{R}.$$

Solution: We construct two vectors lying in the plane. For one, we can take the direction vector of the given line, and for the other, we can take a vector from a known point on the given line to the point P. Thus we let

$$\vec{u} = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} \quad \text{and} \quad \vec{v} = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -4 \\ -1 \end{bmatrix}.$$

Then, since $\vec{u}$ and $\vec{v}$ are nonzero and nonparallel, a vector equation for the plane is

$$\vec{x} = \overrightarrow{OP} + s\vec{u} + t\vec{v} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} + t \begin{bmatrix} 0 \\ -4 \\ -1 \end{bmatrix}, \quad s, t \in \mathbb{R}.$$

Exercise 11 Find parametric equations for the plane given in the previous Example 1.4.5.

We note that for a vector equation for a plane, we do require $\vec{u}$ and $\vec{v}$ to be nonparallel. If $\vec{u}$ and $\vec{v}$ are parallel, say $\vec{u} = c\vec{v}$ for some $c \in \mathbb{R}$, then the vector equation we derive is

$$\vec{x} = \overrightarrow{OP} + s\vec{u} + t\vec{v} = \overrightarrow{OP} + s(c\vec{v}) + t\vec{v} = \overrightarrow{OP} + (sc + t)\vec{v},$$

which is a vector equation for a line through P, not a plane.

Section 1.4 Problems

1.4.1. Find a vector equation for the line L given that

(a) L passes through the point $P(1,-1,3)$ in $\mathbb{R}^3$ and is parallel to the line that passes through the points $A(1,1,2)$ and $B(3,2,-4)$.

(b) L passes through the point $P(1,3)$ in $\mathbb{R}^2$ and is perpendicular to the line with vector equation

$$\vec{x} = \begin{bmatrix} 5 \\ -2 \end{bmatrix} + s\begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad s \in \mathbb{R}.$$

(c) L is given by the equation $x_2 = 3x_1 + 2$ in $\mathbb{R}^2$.

1.4.2. Find a vector equation for the plane T given that

(a) T contains the points $A(1,0,0)$, $B(1,1,0)$ and $C(1,1,1)$.

(b) T contains the two lines with vector equations

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, t \in \mathbb{R} \quad \text{and} \quad \vec{x} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} + s\begin{bmatrix} -4 \\ -2 \\ 6 \end{bmatrix}, s \in \mathbb{R}.$$

1.4.3. Consider the line L in $\mathbb{R}^3$ with vector equation

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} + t\begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}, t \in \mathbb{R}.$$

Determine which of the vector equations below are also vector equations of L.

(a) $\vec{x} = \begin{bmatrix} -3 \\ 6 \\ 0 \end{bmatrix} + t\begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}, t \in \mathbb{R}.$

(b) $\vec{x} = \begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix} + t\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, t \in \mathbb{R}.$

(c) $\vec{x} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} + t\begin{bmatrix} 8 & 12 \\ & -4 \end{bmatrix}, t \in \mathbb{R}.$

(d) $\vec{x} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} + t\begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}, t \in \mathbb{R}.$

1.4.4. Find the point of intersection of the following pairs of lines, or show that no such point exists.

(a) $\vec{x} = \begin{bmatrix} 0 \\ 2 \\ -3 \end{bmatrix} + t\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, t \in \mathbb{R}$ and $\vec{x} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + s\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, s \in \mathbb{R}.$

(b) $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, t \in \mathbb{R}$ and $\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + s\begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}, s \in \mathbb{R}.$

1.4.5. Consider the lines L_1 and L_2 given, respectively, by the vector equations

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, t \in \mathbb{R} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, s \in \mathbb{R}.$$

(a) Find a vector $\vec{v} \in \mathbb{R}^3$ from an aribitrary point on L_1 to an arbitrary point on L_2.

(b) Find s, t so that $\vec{v}$ is orthogonal to the direction vectors of both L_1 and L_2.

(c) For these values of s and t, compute $\|\vec{v}\|$. What does $\|\vec{v}\|$ represent geometrically?

1.4.6. In $\mathbb{R}^3$, consider a line L with vector equation

$$\vec{x} = \begin{bmatrix} 0 \\ 2 \\ -4 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad t \in \mathbb{R},$$

and a plane T with vector equation

$$\vec{x} = \begin{bmatrix} 4 \\ 4 \\ 5 \end{bmatrix} + r \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \quad r, s \in \mathbb{R}.$$

Find the point of intersection of L and T, or show that no such point exists.

1.4.7. Consider two lines in $\mathbb{R}^3$ given by the vector equations

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, t \in \mathbb{R} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 8 \\ 3 \\ 6 \end{bmatrix} + t \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, t \in \mathbb{R}.$$

What follows is an **incorrect** argument that these two lines do not intersect.

We equate the parametric equations of the two lines:

$$\begin{aligned} 1 + t &= 8 + 3t \\ t &= 3 + t \\ 1 + t &= 6 + 2t. \end{aligned}$$

From the second equation, we obtain $3 = 0$*, which is impossible. Thus the two lines do not intersect.*

(a) Where is the error?

(b) Do the lines actually intersect? If so, find the point of intersection.

1.4.8. Find the point of intersection of the two planes in $\mathbb{R}^4$ given by

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + q \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + r \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, q, r \in \mathbb{R} \quad \text{and} \quad \begin{bmatrix} 2 \\ 0 \\ 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, s, t \in \mathbb{R}.$$

Note: In $\mathbb{R}^3$, it should be intuitively clear that if two planes intersect, then their intersection is either a line or plane, but never a single point (a fact that we can verify in Chapter 2). This example shows how different things can be in $\mathbb{R}^4$.

1.5 The Cross Product in $\mathbb{R}^3$

In the previous section, we introduced vector equations for lines and planes. Although our examples were focused in $\mathbb{R}^3$, remember that these equations can be used to describe lines and planes in $\mathbb{R}^n$ as well. Recall that the motivation for the vector equation of a line came from the fact that the equation $ax_1+bx_2=c$ described a line only in $\mathbb{R}^2$. A natural question one may ask is what does the equation $ax_1+bx_2+cx_3=d$ describe in $\mathbb{R}^3$. The beginning of the previous section alluded to the fact that such an equation describes a plane in $\mathbb{R}^3$ (see Figure 1.4.1b). The goal of the next section is to show that the equation $ax_1+bx_2+cx_3=d$ does indeed describe a plane in $\mathbb{R}^3$ and to explain how one derives this equation.

In order to achieve this goal, we will need to define a new operation called the *cross product* which we will examine in this section. This product is only valid[6] in $\mathbb{R}^3$. Whereas the dot product of two vectors in $\mathbb{R}^n$ is a real number, the cross product of two vectors in $\mathbb{R}^3$ is a vector in $\mathbb{R}^3$. The cross product has a rather strange looking definition and satisfies some odd algebraic properties.

Definition 1.5.1
Cross Product in $\mathbb{R}^3$

Let $\vec{x}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$ and $\vec{y}=\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$ be two vectors in $\mathbb{R}^3$. The vector

$$\vec{x}\times\vec{y}=\begin{bmatrix}x_2y_3-y_2x_3\\-(x_1y_3-y_1x_3)\\x_1y_2-y_1x_2\end{bmatrix}\in\mathbb{R}^3$$

is called the **cross product** (or the **vector product**) of $\vec{x}$ and $\vec{y}$.

The formula for $\vec{x}\times\vec{y}$ is quite tedious to remember. Here we give a simpler way. For $a,b,c,d\in\mathbb{R}$, define

$$\begin{vmatrix}a & b\\ c & d\end{vmatrix}=ad-bc$$

so that

$$\vec{x}\times\vec{y}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\times\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}=\begin{bmatrix}\begin{vmatrix}x_2 & y_2\\ x_3 & y_3\end{vmatrix}\\ \\ -\begin{vmatrix}x_1 & y_1\\ x_3 & y_3\end{vmatrix}\\ \\ \begin{vmatrix}x_1 & y_1\\ x_2 & y_2\end{vmatrix}\end{bmatrix}\begin{array}{l}\longleftarrow \text{remove } x_1 \text{ and } y_1\\ \\ \\ \longleftarrow \text{remove } x_2 \text{ and } y_2 \text{ (don't forget the "}-\text{" sign)}\\ \\ \\ \longleftarrow \text{remove } x_3 \text{ and } y_3\end{array}$$

$$=\begin{bmatrix}x_2y_3-y_2x_3\\-(x_1y_3-y_1x_3)\\x_1y_2-y_1x_2\end{bmatrix}.$$

[6] This is not entirely true. There is a cross product in $\mathbb{R}^7$ as well, but it is beyond the scope of this course.

Example 1.5.2 Let $\vec{x} = \begin{bmatrix} 1 \\ 6 \\ 3 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}$. Then

$$\vec{x} \times \vec{y} = \begin{bmatrix} \begin{vmatrix} 6 & 3 \\ 3 & 2 \end{vmatrix} \\ \\ -\begin{vmatrix} 1 & -1 \\ 3 & 2 \end{vmatrix} \\ \\ \begin{vmatrix} 1 & -1 \\ 6 & 3 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 6(2) - 3(3) \\ -\big(1(2) - (-1)(3)\big) \\ 1(3) - (-1)(6) \end{bmatrix} = \begin{bmatrix} 3 \\ -5 \\ 9 \end{bmatrix}.$$

Using the result of Example 1.5.2, we compute

$$\begin{aligned} \vec{x} \cdot (\vec{x} \times \vec{y}) &= 1(3) + 6(-5) + 3(9) = 3 - 30 + 27 = 0 \\ \vec{y} \cdot (\vec{x} \times \vec{y}) &= -1(3) + 3(-5) + 2(9) = -3 - 15 + 18 = 0, \end{aligned}$$

from which we see that $\vec{x} \times \vec{y}$ is orthogonal to both $\vec{x}$ and $\vec{y}$. This is one of the reasons why the cross product will be useful to us: it can produce for us a vector in $\mathbb{R}^3$ that is orthogonal to any two given vectors.

Theorem 1.5.3 Let $\vec{x}, \vec{y} \in \mathbb{R}^3$. Then $\vec{x} \times \vec{y}$ is orthogonal to both $\vec{x}$ and $\vec{y}$.

The proof of Theorem 1.5.3 is simply a matter of showing that the dot products $\vec{x} \cdot (\vec{x} \times \vec{y})$ and $\vec{y} \cdot (\vec{x} \times \vec{y})$ are zero using the expression for $\vec{x} \times \vec{y}$ given in Definition 1.5.1. The algebra is a bit tedious so we will leave it as an exercise. Theorem 1.5.3 will be used in the next section to help us find equations of planes in $\mathbb{R}^3$.

Example 1.5.4 Find a nonzero vector $\vec{n}$ orthogonal to both $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$. Moreover, show that this vector is orthogonal to any linear combination of $\vec{x}$ and $\vec{y}$.

Solution: By Theorem 1.5.3,

$$\vec{n} = \vec{x} \times \vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix}$$

is orthogonal to both $\vec{x}$ and $\vec{y}$. Now for any $s, t \in \mathbb{R}$,

$$\vec{n} \cdot (s\vec{x} + t\vec{y}) = s(\vec{n} \cdot \vec{x}) + t(\vec{n} \cdot \vec{y}) = s(0) + t(0) = 0$$

so $\vec{n} = \vec{x} \times \vec{y}$ is orthogonal to any linear combination of $\vec{x}$ and $\vec{y}$.

Once the cross product of $\vec{x}, \vec{y} \in \mathbb{R}^3$ is computed, we can check that our work is correct by verifying that $(\vec{x} \times \vec{y}\,) \cdot \vec{x} = 0$ and that $(\vec{x} \times \vec{y}\,) \cdot \vec{y} = 0$.

Exercise 12 Check that the vector $\vec{n} = \vec{x} \times \vec{y}$ obtained in Example 1.5.4 is orthogonal to $\vec{x}$ and $\vec{y}$ by computing $\vec{n} \cdot \vec{x}$ and $\vec{n} \cdot \vec{y}$ and showing that they are both equal to 0.

We close off this section with a summary of some of the algebraic properties of the cross product.

Theorem 1.5.5 **(Properties of the Cross Product)**

Let $\vec{x}, \vec{y}, \vec{w} \in \mathbb{R}^3$ and $c \in \mathbb{R}$. Then

(a) $\vec{x} \times \vec{y} \in \mathbb{R}^3$.

(b) $\vec{x} \times \vec{0} = \vec{0} = \vec{0} \times \vec{x}$.

(c) $\vec{x} \times \vec{x} = \vec{0}$.

(d) $\vec{x} \times \vec{y} = -(\vec{y} \times \vec{x})$.

(e) $(c\vec{x}) \times \vec{y} = c(\vec{x} \times \vec{y}) = \vec{x} \times (c\vec{y})$.

(f) $\vec{w} \times (\vec{x} \pm \vec{y}) = (\vec{w} \times \vec{x}) \pm (\vec{w} \times \vec{y})$.

(g) $(\vec{x} \pm \vec{y}) \times \vec{w} = (\vec{x} \times \vec{w}) \pm (\vec{y} \times \vec{w})$.

Proof: We only prove (d). Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ be two vectors in $\mathbb{R}^3$. Then

$$\vec{x} \times \vec{y} = \begin{bmatrix} x_2y_3 - y_2x_3 \\ -(x_1y_3 - y_1x_3) \\ x_1y_2 - y_1x_2 \end{bmatrix} = \begin{bmatrix} -(y_2x_3 - x_2y_3) \\ y_1x_3 - x_1y_3 \\ -(y_1x_2 - x_1y_2) \end{bmatrix} = - \begin{bmatrix} y_2x_3 - x_2y_3 \\ -(y_1x_3 - x_1y_3) \\ y_1x_2 - x_1y_2 \end{bmatrix} = -(\vec{y} \times \vec{x}),$$

as desired. □

Notice that property (d) is a bit unusual. It says that the cross product is not commutative as $\vec{x} \times \vec{y} \neq \vec{y} \times \vec{x}$ in general. The order of $\vec{x}$ and $\vec{y}$ matters. Specifically, changing the order of $\vec{x}$ and $\vec{y}$ in the cross product changes the result by a factor of -1. We indicate this by saying that the cross product is *anti-commutative*. The next exercise exhibits another peculiar property of the cross product.

Exercise 13 Show that the cross product is not associative. That is, find $\vec{x}, \vec{y}, \vec{w} \in \mathbb{R}^3$ such that

$$(\vec{x} \times \vec{y}) \times \vec{w} \neq \vec{x} \times (\vec{y} \times \vec{w}).$$

It follows from Exercise 13 that the expression $\vec{x} \times \vec{y} \times \vec{w}$ is undefined. We must always include brackets to indicate in which order we should evaluate the cross products as changing the order will change the result.

Section 1.5 Problems

1.5.1. Let
$$\vec{x} = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} 2 \\ -6 \\ -4 \end{bmatrix}.$$
Evaluate $\vec{x} \times \vec{y}$, $\vec{x} \times \vec{z}$ and $\vec{y} \times \vec{z}$.

1.5.2. Find a non-zero vector $\vec{z} \in \mathbb{R}^3$ that is orthogonal to both
$$\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} a \\ 2 \\ -a \end{bmatrix}$$
where $a \in \mathbb{R}$ is an arbitrary constant.

1.5.3. Consider a line L containing the point $P(-2, 1, 5)$ with direction vector $\vec{d}$. Find a vector equation for L given that $\vec{d}$ is orthogonal to both $\vec{u} = \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right]$ and $\vec{v} = \left[\begin{smallmatrix} 3 \\ 2 \\ -2 \end{smallmatrix}\right]$.

1.5.4. Let $\vec{x} \in \mathbb{R}^3$.

(a) Prove that $\vec{x} \times \vec{x} = \vec{0}$.

(b) Let $\vec{y} \in \mathbb{R}^3$ be a scalar multiple of $\vec{x}$. Determine $\vec{x} \times \vec{y}$.

1.5.5. Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$. Prove that

(a) $\vec{x} \cdot (\vec{y} \times \vec{z}) = -\vec{y} \cdot (\vec{x} \times \vec{z})$.

(b) $\vec{x} \cdot (\vec{y} \times \vec{z}) = -\vec{z} \cdot (\vec{y} \times \vec{x})$.

1.5.6. Prove Theorem 1.5.3.

1.5.7. Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$. Determine if the following "cancellation law" is true or false:
$$\text{If } \vec{x} \times \vec{y} = \vec{x} \times \vec{z} \text{ then either } \vec{x} = \vec{0} \text{ or } \vec{y} = \vec{z}.$$
If you think this is true, prove it. If you think this is false, give an example showing that it is false.

1.6 The Scalar Equation of Planes in $\mathbb{R}^3$

Given a plane in $\mathbb{R}^3$ and any point P on this plane, there is a unique line through P that is perpendicular to the plane. Let $\vec{n}$ be a direction vector for this line. Then for any point Q on the plane, $\vec{n}$ is orthogonal to $\overrightarrow{PQ}$.

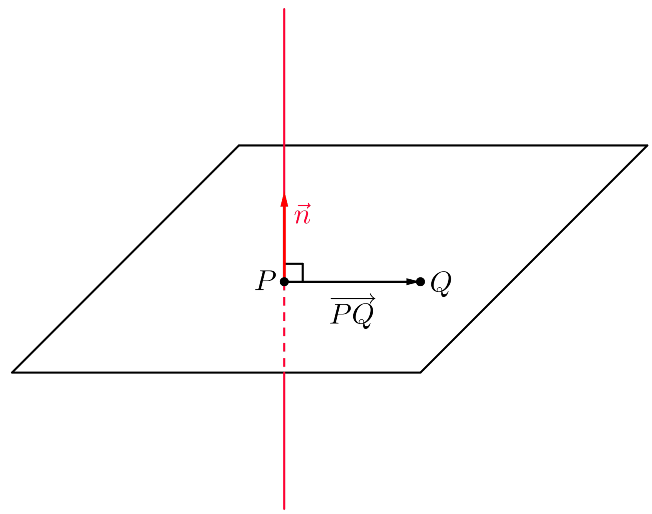


Figure 1.6.1: A line that is perpendicular to a plane.

Definition 1.6.1
Normal Vector for a Plane

A nonzero vector $\vec{n} \in \mathbb{R}^3$ is a **normal vector** for a plane if for any two points P and Q on the plane, $\vec{n}$ is orthogonal to $\overrightarrow{PQ}$.

We note that given a plane in $\mathbb{R}^3$, a normal vector for that plane is not unique as any nonzero scalar multiple of that vector will also be a normal vector for that plane.

Now consider a plane in $\mathbb{R}^3$ with a normal vector

$$\vec{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix},$$

and suppose $P(a, b, c)$ is a given point on this plane. Any point $Q(x_1, x_2, x_3)$ lies on the plane if and only if

$$\begin{aligned} 0 &= \vec{n} \cdot \overrightarrow{PQ} \\ &= \vec{n} \cdot \left(\overrightarrow{OQ} - \overrightarrow{OP}\right) \\ &= \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \cdot \begin{bmatrix} x_1 - a \\ x_2 - b \\ x_3 - c \end{bmatrix} \\ &= n_1(x_1 - a) + n_2(x_2 - b) + n_3(x_3 - c). \end{aligned}$$

That is, $Q(x_1, x_2, x_3)$ will lie on the plane if and only if its coordinates (x_1, x_2, x_3) satisfy the equation

$$n_1(x_1 - a) + n_2(x_2 - b) + n_3(x_3 - c) = 0.$$

Definition 1.6.2

Scalar Equation of a Plane

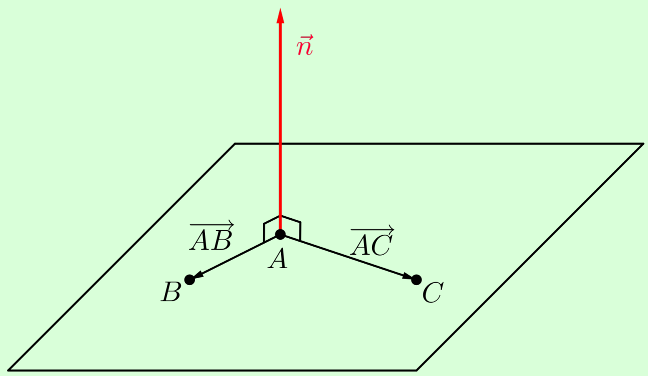
The **scalar equation** of a plane in $\mathbb{R}^3$ with normal vector $\vec{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}$ containing the point $P(a, b, c)$ is given by

$$n_1x_1 + n_2x_2 + n_3x_3 = n_1a + n_2b + n_3c.$$

Example 1.6.3

Find a scalar equation of the plane containing the points $A(3, 1, 2)$, $B(1, 2, 3)$ and $C(-2, 1, 3)$.

Solution: We have three points lying on the plane, so we only need to find a normal vector for the plane.



We compute

$$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

$$\overrightarrow{AC} = \overrightarrow{OC} - \overrightarrow{OA} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix}$$

and notice that $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are nonzero nonparallel vectors in $\mathbb{R}^3$. We compute

$$\vec{n} = \overrightarrow{AB} \times \overrightarrow{AC} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}$$

and recall that the nonzero vector $\vec{n}$ is orthogonal to both $\overrightarrow{AB}$ and $\overrightarrow{AC}$. It follows from Example 1.5.4 that $\vec{n}$ is orthogonal to the entire plane and is thus a normal vector for the plane. Hence, using the point $A(3, 1, 2)$, our scalar equation is

$$1(x_1 - 3) - 3(x_2 - 1) + 5(x_3 - 2) = 0$$

which evaluates to

$$x_1 - 3x_2 + 5x_3 = 10.$$

Exercise 14 Check that the scalar equation given in the previous Example is correct by confirming that the coordinates of the points A, B and C satisfy it.

We make a few remarks about the preceding example here.

- Using the point B or C rather than A to compute the scalar equation would lead to the same scalar equation as is easily verified.
- As the normal vector for the above plane is not unique, neither is the scalar equation. In fact, $2\vec{n}$ is also a normal vector for the plane, and using it instead of $\vec{n}$ would lead to the scalar equation $2x_1 - 6x_2 + 10x_3 = 20$, which is just the scalar equation we found multiplied by a factor of 2.
- From our work above, we see that we can actually compute a vector equation for the plane:
$$\vec{x} = \overrightarrow{OA} + s\overrightarrow{AB} + t\overrightarrow{AC} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} + s\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} + t\begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix}, \quad s,t \in \mathbb{R}$$
for example. In fact, given a vector equation $\vec{x} = \overrightarrow{OP} + s\vec{u} + t\vec{v}$ for a plane in $\mathbb{R}^3$ containing a point P, we can find a normal vector by computing $\vec{n} = \vec{u} \times \vec{v}$.
- Note that in the scalar equation $x_1 - 3x_2 + 5x_3 = 10$, the coefficients on the variables x_1, x_2 and x_3 are exactly the entries in the normal vector we found (see Definition 1.6.2). Thus, if we are given a scalar equation of a different plane, say $3x_1 - 2x_2 + 5x_3 = 72$, we can deduce immediately that $\vec{n} = \begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix}$ is a normal vector for that plane.

Given a plane in $\mathbb{R}^3$, when is it better to use a vector equation and when is it better to use a scalar equation? Consider a plane with scalar equation $4x_1 - x_2 - x_3 = 2$ and vector equation
$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + t\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \quad s,t \in \mathbb{R}.$$
Suppose you are asked if the point $(2, 6, 0)$ lies on this plane. Using the scalar equation $4x_1 - x_2 - x_3 = 2$, we see that $4(2) - 1(6) - 1(0) = 2$ satisfies this equation so we can easily conclude that $(2, 6, 0)$ lies on the plane. However, if we use the vector equation, we must determine if there exist $s, t \in \mathbb{R}$ such that
$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + t\begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 0 \end{bmatrix}$$
which leads to the system of equations
$$\begin{array}{rcrcr} s & + & t & = & 1 \\ 2s & + & t & = & 5 \\ 2s & + & 3t & = & -1 \end{array}$$
With a little work, we can find that the solution to this system[7] is $s = 4$ and $t = -3$ which again guarantees that $(2, 6, 0)$ lies on the plane. It should be clear that using a scalar

[7] We will look at a more efficient technique to solve systems of equations shortly.

equation is preferable here. On the other hand, if you are asked to generate a point that lies on the plane, then using the vector equation, we may select any two values for s and t (say $s = 0$ and $t = 0$) to conclude that the point $(1, 1, 1)$ lies on the plane. It is not too difficult to find a point lying on the plane using the scalar equation either - this will likely be done by choosing two of x_1, x_2, x_3 and then solving for the last, but this does involve a little bit more math. Thus, the scalar equation is preferable when verifying if a given point lies on a plane, and the vector equation is preferable when asked to generate points that lie on the plane.

We have have discussed parallel vectors previously, and we can use this definition to define parallel lines and planes.

Definition 1.6.4
Parallel Lines and Parallel Planes

Two lines in $\mathbb{R}^n$ are **parallel** if their direction vectors are parallel. Two planes in $\mathbb{R}^3$ are **parallel** if their normal vectors are parallel.

Example 1.6.5

The lines in $\mathbb{R}^3$ with vector equations

$$\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} + t \begin{bmatrix} -6 \\ 2 \\ -2 \end{bmatrix}$$

are parallel, since their direction vectors

$$\vec{d}_1 = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \vec{d}_2 = \begin{bmatrix} -6 \\ 2 \\ -2 \end{bmatrix}$$

are parallel. Indeed, $\vec{d}_2 = (-2)\vec{d}_1$.

Exercise 15

Find a vector equation for the line that passes through the point $P(1, 1, 1)$ and is parallel to the lines given in the previous Example.

Example 1.6.6

The planes in $\mathbb{R}^3$ with scalar equations

$$x_1 - x_2 + 3x_3 = 0 \quad \text{and} \quad 3x_1 - 3x_2 + 9x_3 = 4$$

are parallel, since their normal vectors

$$\vec{n}_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{n}_2 = \begin{bmatrix} 3 \\ -3 \\ 9 \end{bmatrix}$$

are parallel. Indeed, $\vec{n}_2 = 3\vec{n}_1$.

Exercise 16

Find a scalar equation for the plane that passes through the point $P(1, 0, 0)$ and is parallel to the planes given in the previous Example.

Section 1.6 Problems

1.6.1. Find a scalar equation of the plane passing through the point $P(2,7,6)$ that is parallel to the plane $2x_1 - 3x_3 = 6$.

1.6.2. Consider a line L with direction vector $\vec{d} \in \mathbb{R}^3$. Find a vector equation for L given that it lies in the plane $x_1 + x_2 + x_3 = 4$, contains the point $P(-2,1,5)$, and that $\vec{d}$ is orthogonal to $\vec{v} = \begin{bmatrix} 3 \\ 2 \\ -2 \end{bmatrix}$.

1.6.3. Consider the plane in $\mathbb{R}^3$ that contains the two lines with vector equations

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix},\ t \in \mathbb{R} \quad \text{and} \quad \vec{x} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} + s\begin{bmatrix} -4 \\ -2 \\ 6 \end{bmatrix},\ s \in \mathbb{R}.$$

Find a scalar equation of this plane.

1.6.4. Consider the points $P(2,1,1)$, $Q(1,2,-1)$, $R(3,2,-1)$ and $S(4,2,3)$. Determine if there is a plane in $\mathbb{R}^3$ that contains all four of these points.

1.6.5. Determine the point(s) of intersection of the line L with the plane T where L has vector equation

$$\vec{x} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \quad t \in \mathbb{R}$$

and T has vector equation

$$\vec{x} = \begin{bmatrix} 4 \\ 1 \\ 6 \end{bmatrix} + s\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + r\begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \quad s, r \in \mathbb{R}.$$

Compare the method you used here to the method used in Problem 1.4.6.

1.7 Projections

Given two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$, we can write $\vec{u} = \vec{u}_1 + \vec{u}_2$ where $\vec{u}_1$ is a scalar multiple of $\vec{v}$ and $\vec{u}_2$ is orthogonal to $\vec{v}$. In physics, this is often done when one wishes to resolve a force into its vertical and horizontal components.

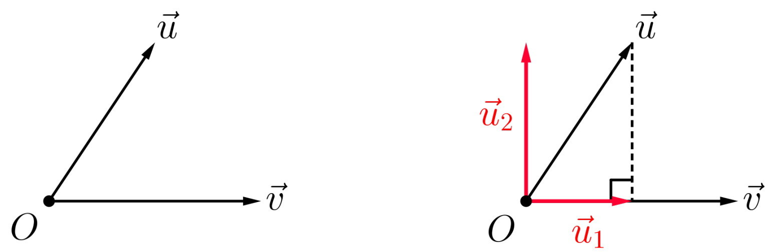


Figure 1.7.1: Decomposing $\vec{u} \in \mathbb{R}^n$ as $\vec{u} = \vec{u}_1 + \vec{u}_2$ where $\vec{u}_1$ is a scalar multiple of $\vec{v}$ and $\vec{u}_2$ is orthogonal to $\vec{v}$.

This is not a new idea. In $\mathbb{R}^2$, we have seen that we can write a vector $\vec{u}$ as a linear combination $\vec{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\vec{e}_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ in a natural way. Figure 1.7.2 shows that we are actually writing a vector $\vec{u} \in \mathbb{R}^2$ as the sum $\vec{u}_1 + \vec{u}_2$ where $\vec{u}_1$ is a scalar multiple of $\vec{v} = \vec{e}_1$ and $\vec{u}_2$ is orthogonal to $\vec{v} = \vec{e}_1$.

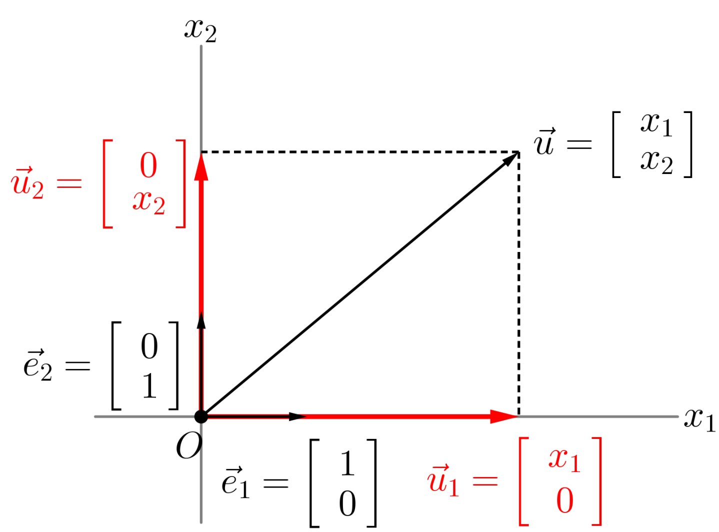


Figure 1.7.2: Writing a vector $\vec{u} = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right] \in \mathbb{R}^2$ as a linear combination of $\vec{e}_1$ and $\vec{e}_2$.

Now for $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$, how do we actually find the vectors $\vec{u}_1$ and $\vec{u}_2$ described above? Let's make a few observations:

$$\begin{aligned} \vec{u} = \vec{u}_1 + \vec{u}_2 &\implies \vec{u}_2 = \vec{u} - \vec{u}_1 && (1.4)\\ \vec{u}_2 \text{ orthogonal to } \vec{v} &\implies \vec{u}_2 \cdot \vec{v} = 0 && (1.5)\\ \vec{u}_1 \text{ a scalar multiple of } \vec{v} &\implies \vec{u}_1 = t\vec{v} \text{ for some } t \in \mathbb{R}. && (1.6) \end{aligned}$$

So if we can determine t, then we can find $\vec{u}_1$ and then finally find $\vec{u}_2$. We have

$$\begin{aligned}
0 &= \vec{u}_2 \cdot \vec{v} && \text{by (1.5)}\\
&= (\vec{u} - \vec{u}_1) \cdot \vec{v} && \text{by (1.4)}\\
&= \vec{u} \cdot \vec{v} - \vec{u}_1 \cdot \vec{v} &&\\
&= \vec{u} \cdot \vec{v} - (t\vec{v}) \cdot \vec{v} && \text{by (1.6)}\\
&= \vec{u} \cdot \vec{v} - t(\vec{v} \cdot \vec{v}) &&\\
&= \vec{u} \cdot \vec{v} - t\|\vec{v}\|^2, &&
\end{aligned}$$

and since $\vec{v} \neq \vec{0}$, we obtain

$$t = \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2},$$

from which we conclude that

$$\vec{u}_1 = \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v} \quad \text{and} \quad \vec{u}_2 = \vec{u} - \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v}.$$

The following definition gives more meaningful names to the vectors $\vec{u}_1$ and $\vec{u}_2$.

Definition 1.7.1

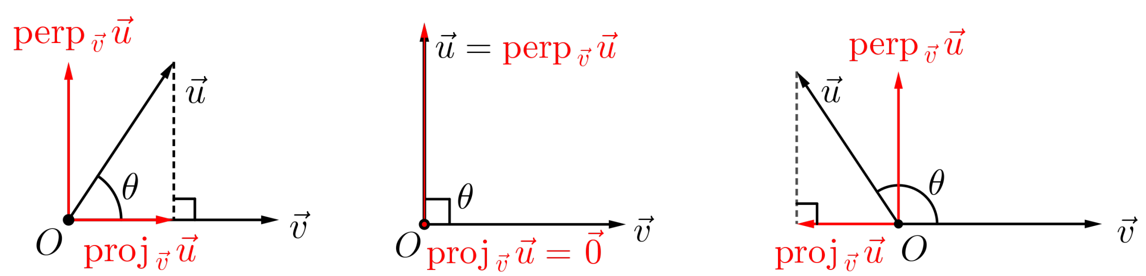
Projection and Perpendicular

Let $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$. The **projection of $\vec{u}$ onto $\vec{v}$** is

$$\operatorname{proj}_{\vec{v}} \vec{u} = \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v}$$

and the **projection of $\vec{u}$ perpendicular to $\vec{v}$** (or the **perpendicular of $\vec{u}$ onto $\vec{v}$**) is

$$\operatorname{perp}_{\vec{v}} \vec{u} = \vec{u} - \operatorname{proj}_{\vec{v}} \vec{u}.$$



(a) The case $0 \leq \theta < \frac{\pi}{2}$, that is, when $\vec{u} \cdot \vec{v} > 0$.

(b) The case $\theta = \frac{\pi}{2}$, that is, when $\vec{u} \cdot \vec{v} = 0$.

(c) The case $\frac{\pi}{2} < \theta \leq \pi$, that is, when $\vec{u} \cdot \vec{v} < 0$.

Figure 1.7.3: Visualizing projections and perpendiculars based on the angle determined by $\vec{u}, \vec{v} \in \mathbb{R}^n$.

Example 1.7.2 Let $\vec{u} = \left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right]$ and $\vec{v} = \left[\begin{smallmatrix}-1\\1\\2\end{smallmatrix}\right]$. Then

$$\operatorname{proj}_{\vec{v}} \vec{u} = \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{-1+2+6}{1+1+4}\begin{bmatrix}-1\\1\\2\end{bmatrix} = \frac{7}{6}\begin{bmatrix}-1\\1\\2\end{bmatrix} = \begin{bmatrix}-7/6\\7/6\\7/3\end{bmatrix}.$$

and

$$\operatorname{perp}_{\vec{v}} \vec{u} = \vec{u} - \operatorname{proj}_{\vec{v}} \vec{u} = \begin{bmatrix}1\\2\\3\end{bmatrix} - \begin{bmatrix}-7/6\\7/6\\7/3\end{bmatrix} = \begin{bmatrix}13/6\\5/6\\2/3\end{bmatrix}.$$

In the previous example, note that

- $\operatorname{proj}_{\vec{v}} \vec{u} = \frac{7}{6}\vec{v}$ which is a scalar multiple of $\vec{v}$,
- $(\operatorname{perp}_{\vec{v}} \vec{u})\cdot\vec{v} = -\frac{13}{6} + \frac{5}{6} + \frac{4}{3} = -\frac{8}{6} + \frac{8}{6} = 0$ so $\operatorname{perp}_{\vec{v}} \vec{u}$ is orthogonal to $\vec{v}$,
- $\operatorname{proj}_{\vec{v}} \vec{u} + \operatorname{perp}_{\vec{v}} \vec{u} = \vec{u}$.

These properties are true in general, and not just for these specific vectors $\vec{u}$ and $\vec{v}$, as you will prove in the exercise below.

Exercise 17 For arbitrary $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$, prove that:

(a) $\operatorname{proj}_{\vec{v}} \vec{u}$ is a scalar multiple of $\vec{v}$.

(b) $\operatorname{perp}_{\vec{v}} \vec{u}$ is orthogonal to $\vec{v}$.

(c) $\operatorname{proj}_{\vec{v}} \vec{u} + \operatorname{perp}_{\vec{v}} \vec{u} = \vec{u}$.

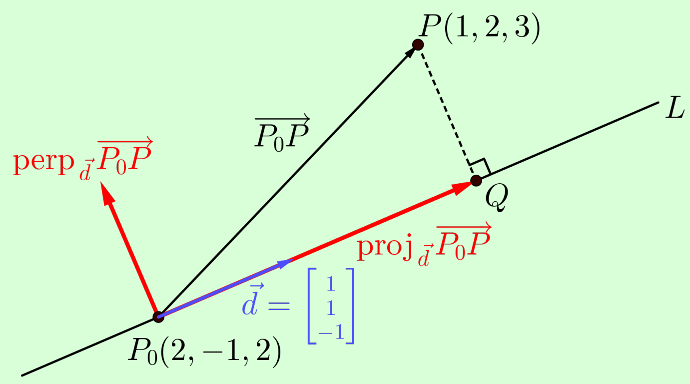
Hints: Definition 1.7.1, Properties of the Norm and Properties of the Dot Product will be helpful here.

1.7.1 Shortest Distances

Given a point P, and the vector equation of a line, we are interested in finding the shortest distance from P to the line, and also the point Q on the line that is closest to P.

Example 1.7.3 Find the shortest distance from the point $P(1,2,3)$ to the line L which passes through the point $P_0(2,-1,2)$ with direction vector $\vec{d} = \left[\begin{smallmatrix}1\\1\\-1\end{smallmatrix}\right]$. Also, find the point Q on L that is closest to P.

Solution: The following illustration can help us visualize the problem. Note that the line L and the point P were plotted arbitrarily, so it is not meant to be accurate. It does however, give us a way to think about the problem geometrically and inform us as to what computations we should do.



We construct the vector from the point P_0 lying on the line to the point P which gives

$$\overrightarrow{P_0P} = \overrightarrow{OP} - \overrightarrow{OP_0} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}.$$

Projecting the vector $\overrightarrow{P_0P}$ onto the direction vector $\vec{d}$ of the line leads to

$$\text{proj}_{\vec{d}}\,\overrightarrow{P_0P} = \frac{\overrightarrow{P_0P}\cdot\vec{d}}{\|\vec{d}\|^2}\vec{d} = \frac{-1+3-1}{1+1+1}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ -1/3 \end{bmatrix}$$

and it follows that

$$\text{perp}_{\vec{d}}\,\overrightarrow{P_0P} = \overrightarrow{P_0P} - \text{proj}_{\vec{d}}\,\overrightarrow{P_0P} = \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/3 \\ 1/3 \\ -1/3 \end{bmatrix} = \begin{bmatrix} -4/3 \\ 8/3 \\ 4/3 \end{bmatrix}.$$

The shortest distance from P to L is thus given by

$$\|\,\text{perp}_{\vec{d}}\,\overrightarrow{P_0P}\| = \frac{1}{3}\sqrt{16+64+16} = \frac{1}{3}\sqrt{16(1+4+1)} = \frac{4}{3}\sqrt{6}.$$

We have two ways to find the point Q since

$$\overrightarrow{OQ} = \overrightarrow{OP_0} + \text{proj}_{\vec{d}}\,\overrightarrow{P_0P} = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1/3 \\ 1/3 \\ -1/3 \end{bmatrix} = \begin{bmatrix} 7/3 \\ -2/3 \\ 5/3 \end{bmatrix}$$

and

$$\overrightarrow{OQ} = \overrightarrow{OP} - \text{perp}_{\vec{d}}\,\overrightarrow{P_0P} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -4/3 \\ 8/3 \\ 4/3 \end{bmatrix} = \begin{bmatrix} 7/3 \\ -2/3 \\ 5/3 \end{bmatrix}.$$

In either case, $Q\left(\frac{7}{3}, -\frac{2}{3}, \frac{5}{3}\right)$ is the point on L closest to P.

We see now we that our illustration in Example 1.7.3 was inaccurate. It seems to suggest that $\operatorname{proj}_{\vec{d}} \overrightarrow{P_0P}$ is approximately $\frac{5}{2}\vec{d}$, but our computations show that $\operatorname{proj}_{\vec{d}} \overrightarrow{P_0P} = \frac{1}{3}\vec{d}$. This is okay, as the illustration was meant only as a guide to inform us as to what computations to perform.

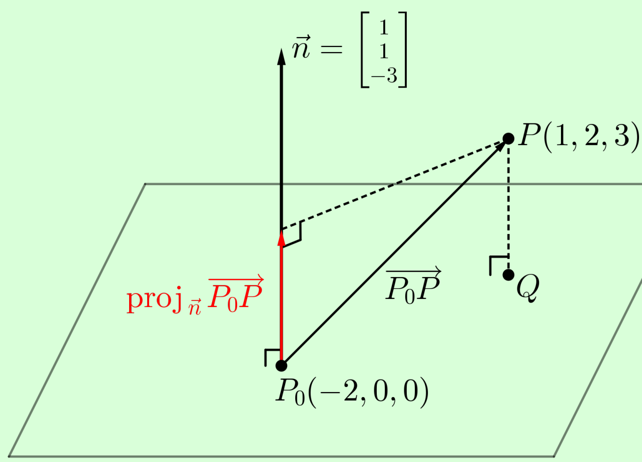
In $\mathbb{R}^3$, given a point P and the scalar equation of a plane, we can also find the shortest distance from P to the plane, as well as the point Q on the plane closest to P.

Example 1.7.4 Find the shortest distance from the point $P(1,2,3)$ to the plane T with equation

$$x_1 + x_2 - 3x_3 = -2.$$

Also, find the point Q on T that is closest to P.

Solution: The accompanying illustration can help us visualize the problem. As in Example 1.7.3, this picture is not meant to be accurate as the point and the line have been plotted arbitrarily, but rather to inform us on what computations we should perform.



We see that $P_0(-2,0,0)$ lies on T since $-2+0-3(0) = -2$. We also have that

$$\vec{n} = \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}$$

is a normal vector for T. Now

$$\overrightarrow{P_0P} = \overrightarrow{OP} - \overrightarrow{OP_0} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}$$

and

$$\operatorname{proj}_{\vec{n}} \overrightarrow{P_0P} = \frac{\overrightarrow{P_0P} \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \frac{3+2-9}{1+1+9}\begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} = -\frac{4}{11}\begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}.$$

The shortest distance from P to T is

$$\|\operatorname{proj}_{\vec{n}} \overrightarrow{P_0P}\| = \left|-\frac{4}{11}\right|\sqrt{1+1+9} = \frac{4\sqrt{11}}{11}.$$

To find Q we have

$$\overrightarrow{OQ} = \overrightarrow{OP} - \text{proj}_{\vec{n}}\,\overrightarrow{P_0P} = \begin{bmatrix}1\\2\\3\end{bmatrix} + \frac{4}{11}\begin{bmatrix}1\\1\\-3\end{bmatrix} = \begin{bmatrix}15/11\\26/11\\21/11\end{bmatrix}$$

so $Q\left(\frac{15}{11}, \frac{26}{11}, \frac{21}{11}\right)$ is the point on T closest to P.

Section 1.7 Problems

1.7.1. Let

$$\vec{x} = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}.$$

(a) Compute $\operatorname{proj}_{\vec{x}} \vec{y}$.

(b) Compute $\operatorname{perp}_{\vec{x}} \vec{y}$.

1.7.2. Consider the point $P(-1, 3, 2)$.

(a) Find the shortest distance from P to the line L with vector equation

$$\vec{x} = \begin{bmatrix} 1 \\ -5 \\ -2 \end{bmatrix} + t \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

and find the point Q on L that is closest to P.

(b) Find the shortest distance from P to the plane T with vector equation

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + r \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} + s \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}, \quad r, s \in \mathbb{R}$$

and find the point Q on T that is closest to P.

1.7.3. Consider the point $P(2, 1, -1)$.

(a) *Without using projections*, find the shortest distance from P to the line L with vector equation

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

and find the point Q on L that is closest to P.

(b) *Without using projections*, find the shortest distance from P to the plane T with scalar equation $x_1 - x_2 + 2x_3 = 3$ and find the point Q on T that is closest to P.

1.7.4. Consider the lines

$$L_1 : \quad \vec{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R},$$

$$L_2 : \quad \vec{x} = \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix} + s \begin{bmatrix} 3 \\ 6 \\ -3 \end{bmatrix}, \quad s \in \mathbb{R}.$$

Find the shortest distance between L_1 and L_2. [**Hint:** you will need to notice something about the direction vectors of L_1 and L_2 and think about how this can be used to find the distance.]

1.7.5. Let L be a line in $\mathbb{R}^3$ with vector equation

$$\vec{x} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

and let T be a plane in $\mathbb{R}^3$ with scalar equation $x_1 - x_2 + 2x_3 = -4$. Using projections, find all points P on L such that the shortest distance from P to T is $2\sqrt{6}$.

1.7.6. Let $\vec{u}, \vec{w} \in \mathbb{R}^n$ with $\vec{w} \neq \vec{0}$ be such that $\vec{u}$ and $\vec{w}$ are not orthogonal. Prove that

$$\operatorname{proj}_{(\operatorname{proj}_{\vec{w}} \vec{u})} \vec{u} = \operatorname{proj}_{\vec{w}} \vec{u}.$$

1.7.7. Let $\vec{u}, \vec{w} \in \mathbb{R}^n$ with $\vec{w} \neq \vec{0}$ and let $k, \ell \in \mathbb{R}$ with $k \neq 0$. Prove that

$$\operatorname{proj}_{k\vec{w}}(\ell \vec{u}) = \ell \operatorname{proj}_{\vec{w}} \vec{u}.$$

1.8 Optional Section: Area and Volume

In this section we will learn how the cross product can be used to compute the area of a parallelogram determined by two vectors in $\mathbb{R}^3$. We will then see how projections, along with some simple geometry, can be used to extend our results to compute the volume of a parallelepiped determined by three vectors in $\mathbb{R}^3$. We begin with the following identity.

Theorem 1.8.1 **(Lagrange Identity)**

Let $\vec{x}, \vec{y} \in \mathbb{R}^3$. Then $\|\vec{x} \times \vec{y}\|^2 = \|\vec{x}\|^2\|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2$.

The proof of the Lagrange Identity is left as an exercise (see Problem 1.8.6). We are now in a position to prove the following result.

Theorem 1.8.2 Let $\vec{x}, \vec{y} \in \mathbb{R}^3$ be nonzero vectors determining an angle θ, where $0 \leq \theta \leq \pi$. Then $\|\vec{x} \times \vec{y}\| = \|\vec{x}\|\|\vec{y}\| \sin\theta$.

Proof: Let $\vec{x}, \vec{y} \in \mathbb{R}^3$ be nonzero vectors. Then by Theorem 1.3.15,

$$\vec{x} \cdot \vec{y} = \|\vec{x}\|\|\vec{y}\| \cos\theta$$

where $0 \leq \theta \leq \pi$. By the Lagrange Identity, we have

$$\begin{aligned}
\|\vec{x} \times \vec{y}\|^2 &= \|\vec{x}\|^2\|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2 \\
&= \|\vec{x}\|^2\|\vec{y}\|^2 - (\|\vec{x}\|\|\vec{y}\| \cos\theta)^2 \\
&= \|\vec{x}\|^2\|\vec{y}\|^2 - \|\vec{x}\|^2\|\vec{y}\|^2 \cos^2\theta \\
&= \|\vec{x}\|^2\|\vec{y}\|^2(1 - \cos^2\theta) \\
&= \|\vec{x}\|^2\|\vec{y}\|^2 \sin^2\theta.
\end{aligned}$$

Since $\sin\theta \geq 0$ for $0 \leq \theta \leq \pi$, we may take square roots to obtain

$$\|\vec{x} \times \vec{y}\| = \|\vec{x}\|\|\vec{y}\| \sin\theta. \qquad \square$$

We now consider the area of the parallelogram determined by vectors $\vec{x}, \vec{y} \in \mathbb{R}^3$. We have the following result.

Theorem 1.8.3 **(Area of a Parallelogram in $\mathbb{R}^3$)**

The area of the parallelogram, P, determined by $\vec{x}, \vec{y} \in \mathbb{R}^3$ is given by

$$\text{area}(P) = \|\vec{x} \times \vec{y}\|.$$

Proof: Let $\vec{x}, \vec{y} \in \mathbb{R}^3$ and let P be the the parallelogram determined by $\vec{x}$ and $\vec{y}$. Consider first the case where neither vector is a scalar multiple of the other, as illustrated in Figure 1.8.1.

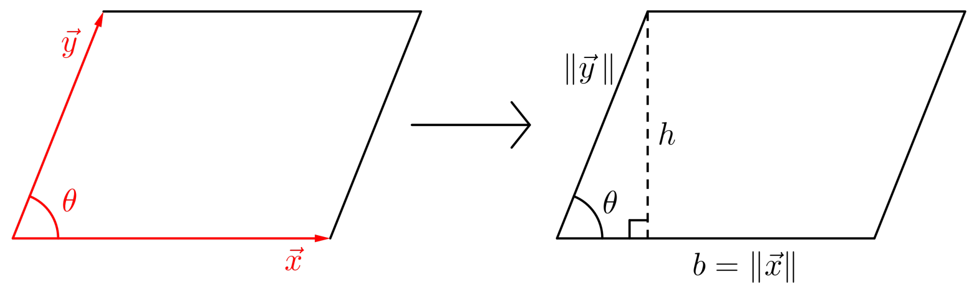


Figure 1.8.1: The parallelogram P determined by $\vec{x}$ and $\vec{y}$.

Define the length of the base of P to be $b = \|\vec{x}\|$. Then the height of P is $h = \|\vec{y}\| \sin\theta$. It follows from Theorem 1.8.2 that

$$\text{area}(P) = bh = \|\vec{x}\|\|\vec{y}\| \sin\theta = \|\vec{x} \times \vec{y}\|.$$

Now consider the case where one of $\vec{x}$ or $\vec{y}$ is a scalar multiple of the other. Then P is a *degenerate parallelogram* (P is a line segment or $P = \{\vec{0}\}$) and it follows that area$(P) = 0$. In this case, we have that $\|\vec{x} \times \vec{y}\| = 0$ (see Problem 1.5.4b). Thus area$(P) = \|\vec{x} \times \vec{y}\|$ for any $\vec{x}, \vec{y} \in \mathbb{R}^3$. $\square$

Example 1.8.4 Let P be the parallelogram determined by $\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$. Find the area of P.

Solution: Since

$$\vec{x} \times \vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -5 \\ 4 \\ 1 \end{bmatrix},$$

we have

$$\text{area}(P) = \|\vec{x} \times \vec{y}\| = \sqrt{25 + 16 + 1} = \sqrt{42}$$

by Theorem 1.8.3.

Exercise 18 Find the area of the parallelogram P determined by the vectors $\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}$ using Theorem 1.8.3. Is the result surprising?

Thus far, our focus has been on two vectors in $\mathbb{R}^3$ and the parallelogram they determine. In a similar fashion, three vectors $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ determined a *parallelepiped* as shown in Figure 1.8.2.

The vectors $\vec{x}, \vec{y}, \vec{z}$ displayed in Figure 1.8.2 are chosen so that none of them is a linear combination of the other two. As a result, any two of these vectors determine a non-degenerate parallelogram and the third vector does not lay in the plane containing this parallelogram.

If $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ are chosen such that at least one of them is a linear combination of the others, then the resulting parallelepiped is a *degenerate parallelepiped* - it will either be planar figure (a "flat" shape), a line segment, or just the set $\{\vec{0}\}$.

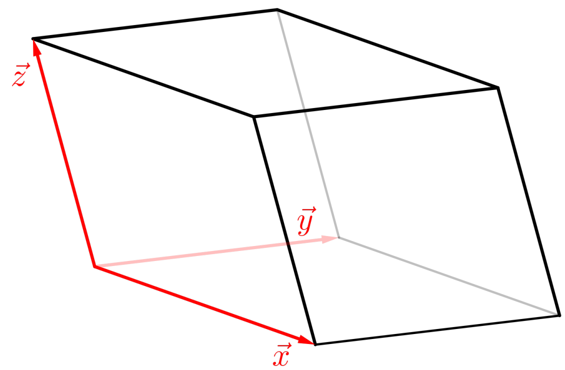


Figure 1.8.2: The parallelepiped determined by the vectors $\vec{x}$, $\vec{y}$ and $\vec{z}$.

Theorem 1.8.5 **(Volume of a Parallelepiped)**

The volume of the parallelepiped, Q, determined by $\vec{z}, \vec{x}, \vec{y} \in \mathbb{R}^3$ is given by

$$\operatorname{vol}(Q) = |\vec{z} \cdot (\vec{x} \times \vec{y})|.$$

Proof: Assume first that no one of $\vec{x}, \vec{y}, \vec{z}$ is a linear combination of the other two. The volume of Q is the product of the area of its base, B and its height, h. Without loss of generality, we assume $\vec{x}$ and $\vec{y}$ determine the base B of Q. By Theorem 1.8.3,

$$\operatorname{area}(B) = \|\vec{x} \times \vec{y}\|.$$

The height, h, of Q is obtained by computing the length of the projection of $\vec{z}$ onto $\vec{x} \times \vec{y}$, that is,

$$h = \|\operatorname{proj}_{\vec{x} \times \vec{y}} \vec{z}\|.$$

This is illustrated in Figure 1.8.3.

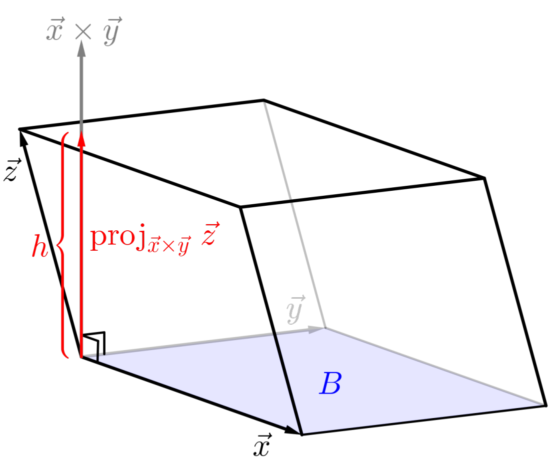


Figure 1.8.3: The parallelepiped Q determined by $\vec{x}$, $\vec{y}$ and $\vec{z}$ with base determined by $\vec{x}$ and $\vec{y}$. The area of the base is $\|\vec{x} \times \vec{y}\|$ and the height is $\|\operatorname{proj}_{\vec{x} \times \vec{y}} \vec{z}\|$.

The volume of Q is given by

$$\begin{aligned}\operatorname{vol}(Q) &= \operatorname{area}(B)\ h \\ &= \|\vec{x}\times\vec{y}\|\|\operatorname{proj}_{\vec{x}\times\vec{y}}\ \vec{z}\| \\ &= \|\vec{x}\times\vec{y}\|\left\|\frac{\vec{z}\cdot(\vec{x}\times\vec{y})}{\|\vec{x}\times\vec{y}\|^2}(\vec{x}\times\vec{y})\right\| \\ &= \|\vec{x}\times\vec{y}\|\frac{|\vec{z}\cdot(\vec{x}\times\vec{y})|}{\|\vec{x}\times\vec{y}\|^2}\|\vec{x}\times\vec{y}\| \\ &= |\vec{z}\cdot(\vec{x}\times\vec{y})|.\end{aligned}$$

We now assume that at least one of $\vec{x}, \vec{y}, \vec{z}$ is a linear combination of the other two. In this case, Q is degenerate parallelepiped and $\operatorname{vol}(Q) = 0$, so we must show that $|\vec{z}\cdot(\vec{x}\times\vec{y})| = 0$.

If $\vec{z}$ is a linear combination of $\vec{x}$ and $\vec{y}$, then it follows from our work in Example 1.5.4 that $\vec{x}\times\vec{y}$ is orthogonal to $\vec{z}$, and from this, we obtain that $|\vec{z}\cdot(\vec{x}\times\vec{y})| = 0$.

If $\vec{z}$ is not a linear combination of $\vec{x}$ and $\vec{y}$, then we have that at least one of $\vec{x}$ and $\vec{y}$ is a scalar multiple of the other. It then follows from our work in Problem 1.5.4b that $\vec{x}\times\vec{y} = \vec{0}$, and hence that $|\vec{z}\cdot(\vec{x}\times\vec{y})| = 0$.

Thus we have shown that $\operatorname{vol}(Q) = |\vec{z}\cdot(\vec{x}\times\vec{y})|$ for any $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$. □

Note that in our derivation of the formula for the volume of the parallelepiped determined by the vectors $\vec{z}$, $\vec{x}$ and $\vec{y}$, we could have chosen any of the six sides of the parallelogram to be the base. Thus, we also have

$$\operatorname{vol}(Q) = |\vec{z}\cdot(\vec{y}\times\vec{x})| = |\vec{x}\cdot(\vec{z}\times\vec{y})| = |\vec{x}\cdot(\vec{y}\times\vec{z})| = |\vec{y}\cdot(\vec{x}\times\vec{z})| = |\vec{y}\cdot(\vec{z}\times\vec{x})|.$$

Example 1.8.6 Let Q be the parallelpiped determined by $\vec{z} = \begin{bmatrix}1\\1\\1\end{bmatrix}$, $\vec{x} = \begin{bmatrix}1\\1\\2\end{bmatrix}$, and $\vec{y} = \begin{bmatrix}1\\2\\-3\end{bmatrix}$. Find the volume of Q.

Solution: We compute

$$\vec{z}\cdot(\vec{x}\times\vec{y}) = \begin{bmatrix}1\\1\\1\end{bmatrix}\cdot\left(\begin{bmatrix}1\\1\\2\end{bmatrix}\times\begin{bmatrix}1\\2\\-3\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\end{bmatrix}\cdot\begin{bmatrix}-7\\5\\1\end{bmatrix} = -7+5+1 = -1$$

so the volume of the parallelepiped determined by $\vec{z}, \vec{x}$ and $\vec{y}$ is

$$\operatorname{vol}(Q) = |\vec{z}\cdot(\vec{x}\times\vec{y})| = |-1| = 1.$$

Exercise 19 Find the volume of the parallelepiped Q determined by the vectors $\vec{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $\vec{e}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$ and $\vec{e}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$. Can you do this without using Theorem 1.8.5?

Section 1.8 Problems

1.8.1. Compute the area of the parallelogram P determined by each of the following pairs of vectors.

(a) $\begin{bmatrix} 4 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix}.$ (b) $\begin{bmatrix} 2 \\ -6 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}.$ (c) $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$

1.8.2. Compute the volume of the parallelpiped Q determined by each of the following 3 vectors.

(a) $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}.$ (b) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}.$ (c) $\begin{bmatrix} 3 \\ -2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -5 \end{bmatrix}.$

1.8.3. Let

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} -1 \\ 2 \\ -2 \end{bmatrix}$$

and let Q be the parallelepiped determined by $\vec{x}$, $\vec{y}$ and $\vec{z}$.

(a) Find the surface area of Q.

(b) Find the volume of Q.

(c) What is the height of Q, given that the base of Q is determined by $\vec{x}$ and $\vec{y}$?

1.8.4. Let $k \in \mathbb{R}$ and let Q be the parallepiped determined by

$$\vec{u} = \begin{bmatrix} 6 \\ 12+4k \\ 2-k \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} 3k \\ 6k+16 \\ k-4 \end{bmatrix}, \quad \text{and} \quad \vec{w} = \begin{bmatrix} 0 \\ -16 \\ 2 \end{bmatrix}.$$

For what value(s) of k is $\mathrm{vol}(Q) = 96$?

1.8.5. Consider the tetrahedron $ABCD$ with vertices $A(1,-1,3)$, $B(1,1,2)$, $C(2,1,-1)$, and $D(2,0,3)$. Let the triangle ABC be the base of the tetrahedron.

(a) Find the area of the base of the tetrahedron.

(b) Find the height of the tetrahedron.

(c) Find the volume of the tetrahedron. [**Hint**: the area of a tetrahedron is one third of the area of its base multiplied by its height.]

(d) Given arbitrary points A, B, C and D in $\mathbb{R}^3$, derive a formula for the volume of the tetrahedron they determine in terms of $\overrightarrow{AB}$, $\overrightarrow{AC}$ and $\overrightarrow{AD}$. You may assume that the tetrahedron is nondegenerate and you may take the triangle determined by A, B and C as the base of the tetrahedron.

1.8.6. Let $\vec{x}, \vec{y} \in \mathbb{R}^3$. Prove the Lagrange Identity, that is, prove that

$$\|\vec{x} \times \vec{y}\|^2 = \|\vec{x}\|^2\|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2.$$

Chapter 2

Systems of Linear Equations

2.1 Introduction and Terminology

In this chapter, we will study systems of linear equations. Such systems are ubiquitous in all scientific fields including engineering. For example, systems of linear equations can arise when one tries to

- balance a chemical reaction,
- determine the current in an electrical network,
- find the stress a steel beam experiences when an external force acts upon it,
- find an optimal solution, such as minimizing the cost of shipping materials from multiple warehouses to multiple worksites.

We have already encountered systems of linear equations in Chapter 1. In Example 1.2.5 we saw a system of equations arise when we verified whether or not a vector could be expressed as a linear combination of a given collection of vectors. In Problem 1.4.4, we again saw systems of equations appearing when we tried to determine if two lines intersected. We offer here one more example that requires us to solve a system of equations.

Example 2.1.1 Find all points that lie on all three planes with scalar equations $2x_1 + x_2 + 9x_3 = 31$, $x_2 + 2x_3 = 8$ and $x_1 + 3x_3 = 10$.

Solution: We are looking for points (x_1, x_2, x_3) that simultaneously satisfy the three equations

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & 9x_3 & = & 31 \\ & & x_2 & + & 2x_3 & = & 8 \\ x_1 & & & + & 3x_3 & = & 10 \end{array}$$

From the second equation, we see that $x_2 = 8 - 2x_3$ and from the third equation, we have that $x_1 = 10 - 3x_3$. Substituting both of these into the first equation gives

$$31 = 2x_1 + x_2 + 9x_3 = 2(10 - 3x_3) + 8 - 2x_3 + 9x_3 = 20 - 6x_3 + 8 - 2x_3 + 9x_3 = 28 + x_3$$

so that $x_3 = 3$. From $x_2 = 8 - 2x_3$, we find that $x_2 = 2$, and from $x_1 = 10 - x_3$ we have $x_1 = 1$. Thus, the only point that lies on all three planes is $(x_1, x_2, x_3) = (1, 2, 3)$.

The methods of elimination and substitution were used to solve the system in Example 2.1.1, but these methods become increasingly cumbersome as the number of equations and variables increase. We thus wish to derive a more systematic method to solve such systems of equations that can more easily extend to handling a large number of equations and variables. This section will be concerned with properly defining a system of linear equations and understanding their basic properties, while the next section will introduce a more efficient way to solve these systems.

Definition 2.1.2

System of Linear Equations

A **linear equation** in n variables is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where $x_1, \ldots, x_n \in \mathbb{R}$ are the **variables** or **unknowns**, $a_1, \ldots, a_n \in \mathbb{R}$ are **coefficients** and $b \in \mathbb{R}$ is the **constant term**. A **system of linear equations** (also called a linear system of equations) is a collection of finitely many linear equations.

Example 2.1.3

A linear equation in three variables is of the form $a_1x_1 + a_2x_2 + a_3x_3 = b$. This is a scalar equation of a plane in $\mathbb{R}^3$, which was discussed in Section 1.6. Note that a single linear equation is still considered a system of linear equations.

Example 2.1.4

The equations

$$3x_1^4 - 2x_2 = 3, \quad x_1 + 2x_1x_3 = 7 \quad \text{and} \quad \sin(x_1) + 3x_2 = 1.$$

are **not** linear equations. The first one is not linear because it contains the term x_1^4. The second one is not linear because it contains the term x_1x_3. The third one is not linear because it contains the term $\sin(x_1)$.

In a linear equation, the variable can only be multiplied by a constant, and no other functions can be applied to them.

Exercise 20

Determine whether the given equation is linear.

(a) $2x_1 + 3x_2 = e^3$.

(b) $2x_1 - 3\dfrac{x_2}{x_3} = 5$.

Example 2.1.5

The system

$$\begin{array}{ccccccccc} 3x_1 & + & 2x_2 & - & x_3 & + & 3x_4 & = & 3 \\ 2x_1 & & & + & x_3 & - & 2x_4 & = & -1 \end{array}$$

is a system of two linear equations in four variables.

More generally, a system of m linear equations in n variables is written as

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

The number a_{ij} is the coefficient of x_j in the ith equation and b_i is the constant term in the ith equation.

Definition 2.1.6

Solution Set of a System of Linear Equations

A vector $\vec{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} \in \mathbb{R}^n$ is a **solution** to a system of m linear equations in n variables if all m equations are satisfied when we set $x_j = s_j$ for $j = 1, \ldots, n$.

The set of all solutions to a system of equations is called the **solution set** of the system.

Example 2.1.7

The vector $\vec{s} = \begin{bmatrix} 3 \\ -5 \\ 0 \end{bmatrix}$ is a solution to the system

$$\begin{array}{ccccccc} 2x_1 & + & x_2 & + & 3x_3 & = & 1 \\ 3x_1 & + & 2x_2 & - & x_3 & = & -1 \\ 5x_1 & + & 3x_2 & + & 2x_3 & = & 0 \end{array}$$

since

$$\begin{array}{ccccccc} 2(3) & + & (-5) & + & 3(0) & = & 1 \\ 3(3) & + & 2(-5) & - & (0) & = & -1 \\ 5(3) & + & 3(-5) & + & 2(0) & = & 0. \end{array}$$

Exercise 21

Check that $\vec{s} = \begin{bmatrix} -4 \\ 6 \\ 1 \end{bmatrix}$ is another solution to the system in Example 2.1.7.

Example 2.1.8

The system of equations in Example 2.1.7 has even more solutions. In fact, every vector of the form

$$\vec{s} = \begin{bmatrix} 3 \\ -5 \\ 0 \end{bmatrix} + t \begin{bmatrix} -7 \\ 11 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R},$$

is a solution to the system. Indeed, if we plug $\vec{s} = \begin{bmatrix} 3-7t \\ -5+11t \\ t \end{bmatrix}$ into the first equation in the system, we get

$$2(3 - 7t) + (-5 + 11t) + 3(t) = 6 - 14t - 5 + 11t + 3t = 1.$$

So the first equation is satisfied. We will leave it to you to check that the second and third equations are satisfied as well.

Later, we will be able to show that the vectors of the form

$$\vec{s} = \begin{bmatrix} 3 \\ -5 \\ 0 \end{bmatrix} + t \begin{bmatrix} -7 \\ 11 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R},$$

make up the entire *solution set* of the system in Example 2.1.7. That is, these vectors are solutions to the system, and there are no other solutions

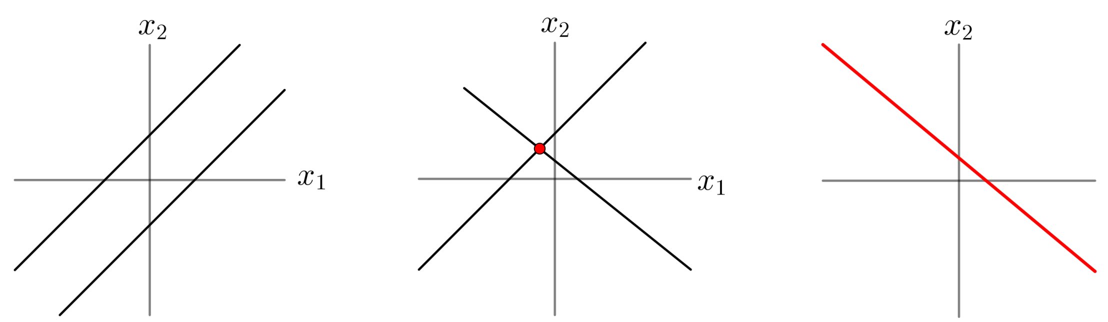
Exercise 22 Complete Example 2.1.8 by showing that the vector $\vec{s} = \begin{bmatrix} 3-7t \\ -5+11t \\ t \end{bmatrix}$ satisfies the second and third equations of the system

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & 3x_3 = & 1 \\ 3x_1 & + & 2x_2 & - & x_3 = & -1 \\ 5x_1 & + & 3x_2 & + & 2x_3 = & 0. \end{array}$$

We now investigate how many solutions a linear system of equations can have. Solving the system of two linear equations in two variables

$$\begin{array}{rcrcl} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & = & b_2 \end{array}$$

can be understood geometrically as finding the points of intersection of the two lines in $\mathbb{R}^2$ with scalar equations $a_{11}x_1 + a_{12}x_2 = b_1$ and $a_{21}x_1 + a_{22}x_2 = b_2$ (where we are assuming that a_{11}, a_{12} are not both zero and that a_{21}, a_{22} are not both zero). Figure 2.1.1 shows the possible number of solutions we may obtain.



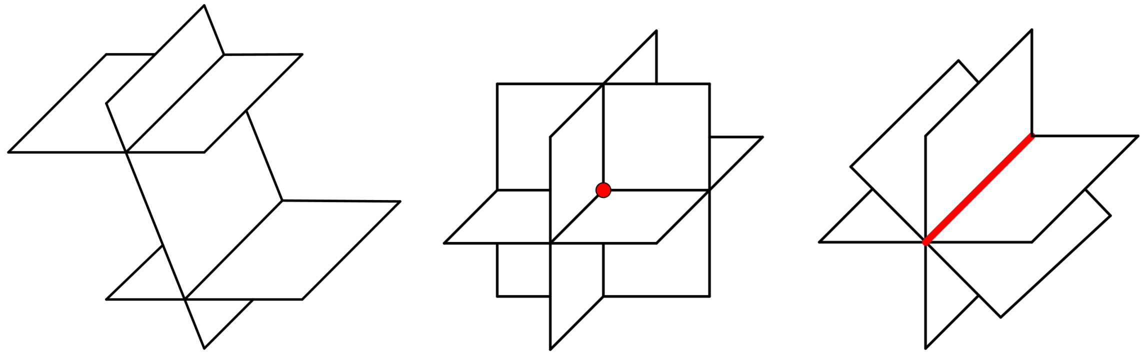
(a) System has no solutions if lines are parallel and distinct.

(b) System has one solution if lines are not parallel.

(c) System has infinitely many solutions if lines are parallel, but not distinct (same lines).

Figure 2.1.1: Number of solutions for a system of two linear equations in two variables which we may view as intersecting two lines in $\mathbb{R}^2$.

We see that a system of two equations in two variables can have no solutions, exactly one solution or infinitely many solutions. Figure 2.1.2 shows a similar situation when we consider a system of three equations in three variables, which we may view geometrically as intersecting three planes in $\mathbb{R}^3$. Indeed we will see that for any linear system of m equations in n variables, we will obtain either no solutions, exactly one solution, or infinitely many solutions. For instance, the system of equations we looked at in Example 2.1.7 and Example 2.1.8 are systems of three equations in three variables that has infinitely many solutions.



(a) No solutions. (b) One solution. (c) Infinitely many solutions.

Figure 2.1.2: Number of solutions for a system of three linear equations in three variables which we may view as intersecting three planes in $\mathbb{R}^3$. Note that there are other ways to arrange these planes to obtain the given number of solutions.

Definition 2.1.9

Consistent, Inconsistent

We call a linear system of equations **consistent** if it has at least one solution. Otherwise, we call the linear system **inconsistent**.

Example 2.1.10

The system

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & 3x_3 & = & 1 \\ 3x_1 & + & 2x_2 & - & x_3 & = & -1 \\ 5x_1 & + & 3x_2 & + & 2x_3 & = & 0 \end{array}$$

from Example 2.1.7 is consistent since we saw that it has $\vec{s} = \begin{bmatrix} 3 \\ -5 \\ 0 \end{bmatrix}$ as a solution.

On the other hand, the system

$$\begin{array}{rcrcr} x_1 & + & 2x_2 & = & -1 \\ x_1 & + & 2x_2 & = & 1 \end{array}$$

is not consistent. This is because the left-sides of the equations are the same but the right-sides are different. Thus, there can be no vector $\vec{s} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ that satisfies both equations simultaneously.

In general, it won't be immediately obvious if a given system of linear equations is consistent or inconsistent. This will be addressed in the next section, where we'll turn our attention to the problem of solving a system of linear equations.

Section 2.1 Problems

2.1.1. Let $a \in \mathbb{R}$. You are told that $\vec{s} = \begin{bmatrix} 5 \\ a \\ 3 \end{bmatrix}$ is a solution to the system

$$\begin{array}{rcrcrl} 2x_1 & + & x_2 & - & x_3 = & 6 \\ x_1 & - & 2x_2 & - & 2x_3 = & 1 \\ -x_1 & + & 12x_2 & + & 8x_3 = & 7 \end{array}$$

Determine a.

2.1.2. Show that, for all $s, t \in \mathbb{R}$,

$$\vec{s} = \begin{bmatrix} 2 \\ 0 \\ 3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 0 \\ 4 \\ 1 \end{bmatrix}$$

is a solution to the system

$$\begin{array}{rcrcrcrl} 2x_1 & + & 4x_2 & + & x_3 & - & 6x_4 = & 7 \\ 4x_1 & + & 8x_2 & - & 3x_3 & + & 8x_4 = & -1 \\ -3x_1 & - & 6x_2 & + & 2x_3 & - & 5x_4 = & 0 \\ x_1 & + & 2x_2 & + & x_3 & - & 5x_4 = & 5 \end{array}$$

2.1.3. Show that the system of equations

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

is consistent no matter what the values of $a_{11}, \ldots, a_{mn}$ are.

2.1.4. Give values $a, b \in \mathbb{R}$ that guarantee that the system

$$\begin{array}{rcrcr} ax_1 & + & 2x_2 & = & 1 \\ x_1 & + & bx_2 & = & -1 \end{array}$$

is inconsistent.

2.2 Solving Systems of Linear Equations

We now present a more systematic way to solve systems of linear equations. We begin by solving a simple system of two linear equations in two variables.

Example 2.2.1 Solve the linear system

$$\begin{array}{rcrcr} x_1 & + & 3x_2 & = & -1 \\ x_1 & + & x_2 & = & 3 \end{array}.$$

Solution: To begin, we will eliminate x_1 in the second equation by subtracting the first equation from the second:

$$\begin{array}{rcrcr} x_1 & + & 3x_2 & = & -1 \\ x_1 & + & x_2 & = & 3 \end{array} \longrightarrow \left(\begin{array}{c} \text{Subtract the first} \\ \text{equation from the second} \end{array} \right) \longrightarrow \begin{array}{rcrcr} x_1 & + & 3x_2 & = & -1 \\ & & -2x_2 & = & 4 \end{array}.$$

Next, we multiply the second equation by a factor of $-\frac{1}{2}$:

$$\begin{array}{rcrcr} x_1 & + & 3x_2 & = & -1 \\ & & -2x_2 & = & 4 \end{array} \longrightarrow \left(\begin{array}{c} \text{Multiply second} \\ \text{equation by } -\frac{1}{2} \end{array} \right) \longrightarrow \begin{array}{rcrcr} x_1 & + & 3x_2 & = & -1 \\ & & x_2 & = & -2 \end{array}.$$

Finally we eliminate x_2 from the first equation by subtracting the second equation from the first equation three times:

$$\begin{array}{rcrcr} x_1 & + & 3x_2 & = & -1 \\ & & x_2 & = & -2 \end{array} \longrightarrow \left(\begin{array}{c} \text{Subtract 3 times the second} \\ \text{equation from the first} \end{array} \right) \longrightarrow \begin{array}{rcrcr} x_1 & & & = & 5 \\ & & x_2 & = & -2 \end{array}.$$

From here, we conclude that the given system is consistent with $x_1 = 5$ and $x_2 = -2$. Thus our solution is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}.$$

Notice that when we write a system of equations, we always list the variables in order from left to right, and that when we solve a system of equations, we are ultimately concerned with the coefficients and constant terms. Thus, we can write the above systems of equations and the subsequent operations we used to solve the system more compactly:

$$\left[\begin{array}{cc|c} 1 & 3 & -1 \\ 1 & 1 & 3 \end{array}\right] \xrightarrow[R_2 - R_1]{} \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & -2 & 4 \end{array}\right] \xrightarrow[-\frac{1}{2}R_2]{} \left[\begin{array}{cc|c} 1 & 3 & -1 \\ 0 & 1 & -2 \end{array}\right] \xrightarrow{R_1 - 3R_2} \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -2 \end{array}\right]$$

so

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

as above. We call

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

the *coefficient matrix*[1] of the linear system, which is often denoted by the letter A. The vector

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

[1] A **matrix** will be formally defined in Chapter 3 - for now, we view them as rectangular arrays of numbers used to represent systems of linear equations.

is the *constant matrix* (or constant vector) of the linear system and will be denoted by the letter $\vec{b}$. Finally

$$\left[\begin{array}{cc|c} 1 & 3 & -1 \\ 1 & 1 & 3 \end{array}\right]$$

is the *augmented matrix* of the linear system, and will be denoted by $\left[A \mid \vec{b}\right]$. This is generalized in the following definition.

Definition 2.2.2

Coefficient Matrix,Augmented Matrix, Constant Vector

For the system of linear equations

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

the **coefficient matrix** is

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

the **constant vector** is

$$\vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

and the **augmented matrix** is

$$\left[A \mid \vec{b}\right] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

Exercise 23

Write down the coefficient matrix A and constant vector $\vec{b}$ of the system

$$\begin{array}{cccccc} 2x_1 & + & x_2 & + & 3x_3 = & 1 \\ 3x_1 & + & 2x_2 & - & x_3 = & -1 \\ 5x_1 & + & 3x_2 & + & 2x_3 = & 0. \end{array}$$

From the discussion immediately following Example 2.2.1, we see that by taking the augmented matrix of a linear system of equations, we can "reduce" it to an augmented matrix of a simpler system from which we can "read off" the solution. Notice that by writing things in this way, we are simply suppressing the variables (since we know x_1 is always the first variable and x_2 is always the second variable), and treating the equations as rows of the augmented matrix. Thus, the operation $R_2 - R_1$ written to the right of the second row of an augmented matrix means that we are subtracting the first row from the second to obtain a new second row which will appear in the next augmented matrix. The following definition summarizes the operations we are allowed to perform to an augmented matrix.

Definition 2.2.3

Elementary Row Operations

We are allowed to perform the following **Elementary Row Operations** (EROs) to the augmented matrix of a linear system of equations:

- Swap two rows.
- Add a scalar multiple of one row to another.
- Multiply any row by a *nonzero* scalar.

We say that two systems are *equivalent* if they have the same solution set. A system derived from a given system by performing elementary row operations to its augmented matrix will be equivalent to the given system. Thus elementary row operations allow us to reduce a complicated system to one that is easier to solve. In the previous example, we applied elementary row operations to arrive at

$$\left[\begin{array}{cc|c} 1 & 3 & -1 \\ 1 & 1 & 3 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -2 \end{array}\right].$$

Consequently, the systems represented by the two augmented matrices above, namely

$$\begin{array}{rcrcr} x_1 & + & 3x_2 & = & -1 \\ x_1 & + & x_2 & = & 3 \end{array} \quad \text{and} \quad \begin{array}{rcrcr} x_1 & & & = & 5 \\ & & x_2 & = & -2 \end{array},$$

must have the same solution set. Clearly, the second system is easier to solve as we can simply read off the solution.

Let's return to the system of linear equations from Example 2.1.1. We will attempt to solve the system by performing elementary row operations on its augmented matrix.

Example 2.2.4

Solve the linear system of equations

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & 9x_3 & = & 31 \\ & & x_2 & + & 2x_3 & = & 8 \\ x_1 & & & + & 3x_3 & = & 10 \end{array}$$

Solution: To solve this system, we perform elementary row operations to the augmented matrix:

$$\left[\begin{array}{ccc|c} 2 & 1 & 9 & 31 \\ 0 & 1 & 2 & 8 \\ 1 & 0 & 3 & 10 \end{array}\right] \underset{R_1 \leftrightarrow R_3}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 0 & 3 & 10 \\ 0 & 1 & 2 & 8 \\ 2 & 1 & 9 & 31 \end{array}\right] \underset{R_3 - 2R_1}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 0 & 3 & 10 \\ 0 & 1 & 2 & 8 \\ 0 & 1 & 3 & 11 \end{array}\right] \underset{R_3 - R_2}{\longrightarrow}$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 3 & 10 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{array}\right] \underset{\substack{R_1 - 3R_3 \\ R_2 - 2R_3}}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

We thus have $x_1 = 1$, $x_2 = 2$ and $x_3 = 3$. Thus our solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

It is likely unclear which elementary row operations one should perform on an augmented matrix in order to solve a linear system of equations. Note that in the two examples above, we eventually arrived at

$$\left[\begin{array}{cc|c} 1 & 3 & -1 \\ 1 & 1 & 3 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -2 \end{array}\right]$$

and

$$\left[\begin{array}{ccc|c} 2 & 1 & 9 & 31 \\ 0 & 1 & 2 & 8 \\ 1 & 0 & 3 & 10 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

The augmented matrices on the right represent simpler systems of linear equations whose solutions can be read off immediately. It would be ideal if we could choose our elementary row operations in order to get to augmented matrices that have the same "form" as these two augmented matrices.

Definition 2.2.5

Row Echelon Form, Reduced Row Echelon Form, Leading Entry, Leading One

The first nonzero entry in each row of a matrix is called a **leading entry** (or a **pivot**). A matrix is in **Row Echelon Form** (REF) if

(a) all rows whose entries are all zero appear below all rows that contain nonzero entries,

(b) each leading entry is to the right of the leading entries above it.

A matrix is in **Reduced Row Echelon Form** (RREF) if it is in REF and additionally

(c) each leading entry is a 1, called a **leading one**,

(d) each leading one is the only nonzero entry in its column.

Note that by definition, if a matrix is in RREF, then it is in REF.

Example 2.2.6

The matrix

$$A = \begin{bmatrix} 0 & 0 & 3 & 3 \\ 5 & 2 & -1 & 2 \end{bmatrix}$$

is not in REF. Its leading entries have been circled. The leading entry 5 appears to the left of the leading entry 3 that is above it.

The matrix

$$B = \begin{bmatrix} 5 & 2 & -1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 3 \end{bmatrix}$$

is not in REF. Although the bottom leading entry is to the right of the top leading entry, there is a zero row that appears between them.

The matrix

$$C = \begin{bmatrix} 5 & 2 & -1 & 2 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in REF. Notice how each leading entry is to the right of the leading entries above it, and that any zero rows appear below all nonzero rows.

Example 2.2.7 The matrix

$$A = \begin{bmatrix} 1 & 4 & 3 & 5 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in REF. Its leading entries have been circled. Since the leading entry in the second row is a -2 and not a 1, the matrix A is not in RREF.

The matrix

$$B = \begin{bmatrix} 1 & 4 & 3 & 5 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in REF but not in RREF. Even though each leading entry is a 1, there is a nonzero entry other than the leading 1 (namely, a 4) in the second column. In RREF, if a column contains a leading entry, then that leading entry must be the only non-zero entry in the column.

Finally, the matrix

$$B = \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in RREF (and REF).

Exercise 24 Determine which of the following matrices is in REF, RREF or neither:

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

$$D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

When row reducing the augmented matrix of a linear system of equations, we aim first to reduce the augmented matrix to REF. Once we have reached an REF form, we continue using elementary row operations until we reach RREF where we can simply read off the solution.

Recalling Example 2.2.4, we rewrite the steps taken to row reduce the augmented matrix of the system and circle the leading entries:

$$\left[\begin{array}{ccc|c} 2 & 1 & 9 & 31 \\ 0 & 1 & 2 & 8 \\ 1 & 0 & 3 & 10 \end{array}\right] \xrightarrow[R_1 \leftrightarrow R_3]{} \left[\begin{array}{ccc|c} 1 & 0 & 3 & 10 \\ 0 & 1 & 2 & 8 \\ 2 & 1 & 9 & 31 \end{array}\right] \xrightarrow[R_3 - 2R_1]{} \left[\begin{array}{ccc|c} 1 & 0 & 3 & 10 \\ 0 & 1 & 2 & 8 \\ 0 & 1 & 3 & 11 \end{array}\right] \xrightarrow[R_3 - R_2]{}$$

$$\underbrace{\left[\begin{array}{ccc|c} 1 & 0 & 3 & 10 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{array}\right]}_{\text{REF}} \xrightarrow[\substack{R_1-3R_3 \\ R_2-2R_3}]{} \underbrace{\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]}_{\text{REF and RREF}}$$

We point out here that if a matrix has at least one nonzero entry, then it will have infinitely many REFs, but the RREF of any matrix is unique.

Example 2.2.8 Solve the linear system of equations

$$\begin{array}{rcrcrcr} 3x_1 & + & x_2 & & & = & 10 \\ 2x_1 & + & x_2 & + & x_3 & = & 6 \\ -3x_1 & + & 4x_2 & + & 15x_3 & = & -20 \end{array}$$

Solution: We use elementary row operations to carry the augmented matrix of the system to RREF.

$$\left[\begin{array}{ccc|c} 3 & 1 & 0 & 10 \\ 2 & 1 & 1 & 6 \\ -3 & 4 & 15 & -20 \end{array}\right] \xrightarrow{R_1-R_2} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 2 & 1 & 1 & 6 \\ -3 & 4 & 15 & -20 \end{array}\right] \xrightarrow[\substack{R_2-2R_1 \\ R_3+3R_1}]{} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 0 & 1 & 3 & -2 \\ 0 & 4 & 12 & -8 \end{array}\right] \xrightarrow[R_3-4R_2]{} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

If we write out the resulting system, we have

$$\begin{array}{rcrcr} x_1 & & - & x_3 & = & 4 \\ & x_2 & + & 3x_3 & = & -2 \\ & & & 0 & = & 0 \end{array}$$

The last equation is clearly always true, and from the first two equations, we can solve for x_1 and x_2 respectively to obtain

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

We see that there is no restriction on x_3, so we let $x_3 = t \in \mathbb{R}$. Thus our solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4+t \\ -2-3t \\ t \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

Geometrically, we view solving the above system of equations as finding those points in $\mathbb{R}^3$ that lie on the three planes $3x_1 + x_2 = 10$, $2x_1 + x_2 + x_3 = 6$ and $-3x_1 + 4x_2 + 15x_3 = -20$. Notice that the solution we obtained

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

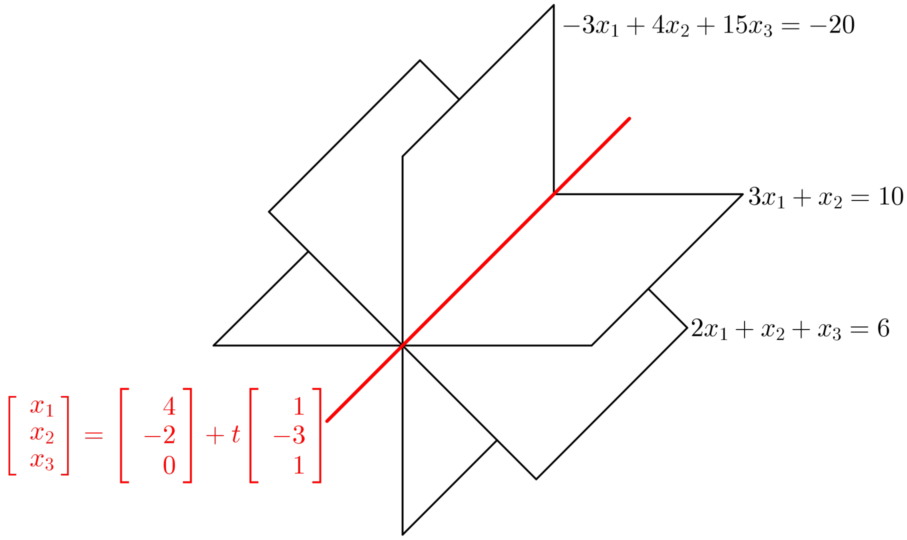


Figure 2.2.1: The intersection of the three planes in $\mathbb{R}^3$ is a line. Note that the planes may not be arranged exactly as shown.

is the vector equation of a line in $\mathbb{R}^3$. Hence we see that the three planes intersect in a line, and we have found a vector equation for that line. See Figure 2.2.1.

That our solution was a line in $\mathbb{R}^3$ was a direct consequence of the fact that there were no restrictions on the variable x_3 and that as a result, our solutions for x_1 and x_2 depended on x_3. This motivates the following definition.

Definition 2.2.9

Leading Variable and Free Variable

Consider a consistent system of equations with augmented matrix $\left[A \mid \vec{b}\right]$, and let $\left[R \mid \vec{c}\right]$ be any REF of $\left[A \mid \vec{b}\right]$. If the jth column of R has a leading entry in it, then the variable x_j is called a **leading variable**. If the jth column of R does not have a leading entry, then x_j is called a **free variable**.

In our last example,

$$\left[\begin{array}{ccc|c} 3 & 1 & 0 & 10 \\ 2 & 1 & 1 & 6 \\ -3 & 4 & 15 & -20 \end{array}\right] \xrightarrow{R_1 - R_2} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 2 & 1 & 1 & 6 \\ -3 & 4 & 15 & -20 \end{array}\right] \xrightarrow[R_3+3R_1]{R_2-2R_1}$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 0 & 1 & 3 & -2 \\ 0 & 4 & 12 & -8 \end{array}\right] \xrightarrow[R_3-4R_2]{} \underbrace{\left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right]}_{\text{REF and RREF}}$$

With $R = \left[\begin{smallmatrix} 1 & 0 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{smallmatrix}\right]$ being an RREF (and thus an REF) of the coefficient matrix of the linear system of equations, we see that R has leading entries (leading ones, in fact) in the first and second columns only. Thus Definition 2.2.9 states that x_1 and x_2 are leading variables while x_3 is a free variable.

When solving a consistent system, if there are free variables, then each free variable is assigned a different parameter, and the leading variables are then solved for in terms of the parameters. The existence of a free variable guarantees that there will be infinitely many solutions to the linear system of equations.

Example 2.2.10 Solve the linear system of equations

$$\begin{array}{rcrcrcrcr} x_1 & + & 6x_2 & & & - & x_4 & = & -1 \\ & & & & x_3 & + & 2x_4 & = & 7. \end{array}$$

Solution: The augmented matrix for this system of linear equations

$$\left[\begin{array}{cccc|c} 1 & 6 & 0 & -1 & -1 \\ 0 & 0 & 1 & 2 & 7 \end{array}\right]$$

is already in RREF. The leading entries are in the first and third columns, so x_1 and x_3 are leading variables while x_2 and x_4 are free variables. We will assign x_2 and x_4 *different* parameters. We have

$$\begin{array}{rcc} x_1 & = & -1-6s+t \\ x_2 & = & s \\ x_3 & = & 7-2t \\ x_4 & = & t \end{array}, \quad s,t \in \mathbb{R}$$

so our solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 7 \\ 0 \end{bmatrix} + s\begin{bmatrix} -6 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \quad s,t \in \mathbb{R}$$

which we recognize as the vector equation of a plane in $\mathbb{R}^4$.

Example 2.2.11 Solve the linear system of equations

$$\begin{array}{rcrcrcr} 2x_1 & + & 12x_2 & - & 8x_3 & = & -4 \\ 2x_1 & + & 13x_2 & - & 6x_3 & = & -5 \\ -2x_1 & - & 14x_2 & + & 4x_3 & = & 7. \end{array}$$

Solution: We have

$$\left[\begin{array}{ccc|c} 2 & 12 & -8 & -4 \\ 2 & 13 & -6 & -5 \\ -2 & -14 & 4 & 7 \end{array}\right] \underset{\substack{R_2-R_1 \\ R_3+R_1}}{\longrightarrow} \left[\begin{array}{ccc|c} 2 & 12 & -8 & -4 \\ 0 & 1 & 2 & -1 \\ 0 & -2 & -4 & 3 \end{array}\right] \underset{R_3+2R_2}{\longrightarrow} \left[\begin{array}{ccc|c} 2 & 12 & -8 & -4 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

The resulting system is

$$\begin{array}{rcrcrcr} 2x_1 & + & 12x_2 & - & 8x_3 & = & -4 \\ & & x_2 & + & 2x_3 & = & -1 \\ & & & & 0 & = & 1. \end{array}$$

Clearly, the last equation can never be satisfied for any $x_1, x_2, x_3 \in \mathbb{R}$. Hence our system is inconsistent, that is, it has no solution.

Geometrically, we see that the three planes $2x_1 + 12x_2 - 8x_3 = -4$, $2x_1 + 13x_2 - 6x_3 = -5$ and $-2x_1 - 14x_2 + 4x_3 = 7$ of Example 2.2.11 have no point in common. Notice that no two of these planes are parallel so the planes are arranged similarly to what is depicted in Figure 2.2.2.

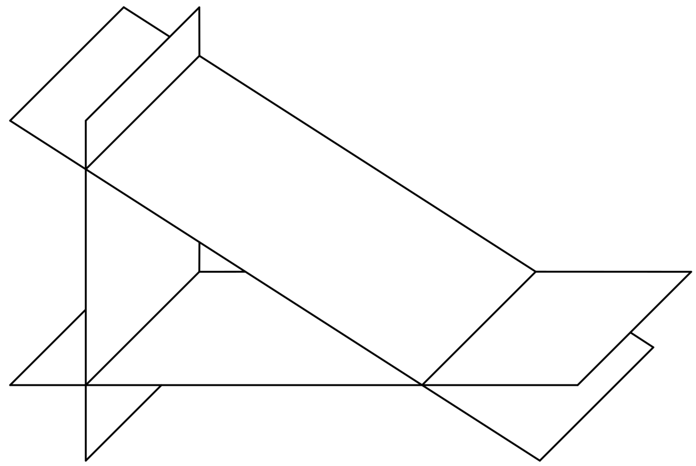


Figure 2.2.2: Three nonparallel planes that have no common point of intersection.

Keeping track of our leading entries in Example 2.2.11, we see

$$\left[\begin{array}{ccc|c} 2 & 12 & -8 & -4 \\ 2 & 13 & -6 & -5 \\ -2 & -14 & 4 & 7 \end{array}\right] \xrightarrow[R_3+R_1]{R_2-R_1} \left[\begin{array}{ccc|c} 2 & 12 & -8 & -4 \\ 0 & 1 & 2 & -1 \\ 0 & -2 & -4 & 3 \end{array}\right] \xrightarrow[R_3+2R_2]{} \underbrace{\left[\begin{array}{ccc|c} 2 & 12 & -8 & -4 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right]}_{\text{REF (but not RREF)}}.$$

If row reducing an augmented matrix reveals a row of the form

$$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & c \end{array}\right]$$

with $c \neq 0$, then the system is inconsistent. Thus, there is no need to continue row operations in this case. Note that in a row of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & c \end{array}\right]$ with $c \neq 0$, the entry c is a leading entry. Thus, a leading entry appearing in the last column of an augmented matrix indicates that the system of linear equations is inconsistent.

Section 2.2 Problems

2.2.1. Find the solutions (if they exist) to the following systems of linear equations by row reducing the augmented matrix. Clearly state the Elementary Row Operations (EROs) you use. If the system is consistent, carry the augmented matrix to reduced row echelon form. If the system is inconsistent, clearly justify why.

(a) $$\begin{array}{rcrcr} 6x_1 & + & 7x_2 & = & 11 \\ 3x_1 & + & 2x_2 & = & 4 \end{array}.$$

(b) $$\begin{array}{rcrcrcrcr} 2x_1 & - & 2x_2 & + & x_3 & + & 15x_4 & = & 6 \\ 3x_1 & - & 3x_2 & + & x_3 & + & 21x_4 & = & 8 \end{array}.$$

(c) $$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 3x_3 & = & 1 \\ x_1 & + & 3x_2 & + & x_3 & = & -1 \\ 2x_1 & + & 2x_2 & + & 10x_3 & = & 8 \end{array}.$$

(d) $$\begin{array}{rcrcrcr} x_1 & + & 3x_2 & + & 10x_3 & = & -18 \\ 2x_1 & + & 6x_2 & + & 19x_3 & = & -34 \\ -x_1 & - & 2x_2 & - & 6x_3 & = & 11 \end{array}.$$

(e) $$\begin{array}{rcrcrcrcrcr} 2x_1 & - & 4x_2 & + & x_3 & + & 2x_4 & + & 3x_5 & = & -2 \\ x_1 & - & 2x_2 & + & x_3 & + & 2x_4 & + & x_5 & = & 1 \\ 3x_1 & - & 6x_2 & + & 3x_3 & + & 3x_4 & + & 6x_5 & = & 3 \end{array}.$$

(f) $$\begin{array}{rcrcrcrcrcr} x_1 & + & 2x_2 & + & x_3 & + & 3x_4 & + & 4x_5 & = & 0 \\ x_1 & + & 2x_2 & + & 2x_3 & + & 5x_4 & + & 5x_5 & = & 0 \\ 2x_1 & + & 4x_2 & + & x_3 & + & 4x_4 & + & 7x_5 & = & 0 \end{array}.$$

2.2.2. Consider the system of linear equations

$$\begin{array}{rcrcrcr} 2x_1 & + & 3x_2 & + & x_3 & = & 1 \\ 2x_1 & + & x_2 & - & x_3 & = & 3 \end{array}.$$

(a) By interpreting the system as giving the solution to a geometry problem, explain why there will either be no solutions or infinitely many solutions.

(b) By thinking more carefully about the geometric problem, determine whether this system will have no solution or infinitely many solutions.

(c) Solve the system by row reducing its augmented matrix. Interpret your result geometrically.

2.2.3. Determine a vector equation for each of the following planes in $\mathbb{R}^3$.

(a) $x_1 + x_2 + x_3 = 2$. (b) $2x_1 + 3x_2 - 5x_3 = -2$. (c) $x_3 = 1$.

[**Hint**: The scalar equation of a plane in $\mathbb{R}^3$ can be thought of as a system of one linear equation.]

2.2.4. In $\mathbb{R}^2$, consider a parabola of the form $y = a_0 + a_1x + a_2x^2$ where $a_0, a_1, a_2 \in \mathbb{R}$ with $a_2 \neq 0$.

(a) Find the equation of such a parabola passing through the points $P(-1, 19)$, $Q(1, 1)$ and $R(2, 1)$.

(b) Show that there is no such parabola passing through the points $P(0, -1)$, $Q(1, 1)$ and $R(3, 5)$.

(c) Show that there are infinitely many such parabolas passing through the points $P(-2, 2)$ and $Q(2, 3)$.

2.2.5. If possible, express

(a) $\begin{bmatrix}-2\\-3\\-6\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\2\\1\end{bmatrix}$, $\begin{bmatrix}3\\1\\2\end{bmatrix}$ and $\begin{bmatrix}1\\-1\\-1\end{bmatrix}$.

(b) $\begin{bmatrix}5\\0\\4\end{bmatrix}$ as a linear combination of $\begin{bmatrix}2\\1\\3\end{bmatrix}$, $\begin{bmatrix}6\\-2\\2\end{bmatrix}$ and $\begin{bmatrix}0\\-5\\-7\end{bmatrix}$.

(c) $\begin{bmatrix}1\\1\\1\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\2\\7\end{bmatrix}$, $\begin{bmatrix}-1\\1\\3\end{bmatrix}$ and $\begin{bmatrix}2\\1\\4\end{bmatrix}$.

2.2.6. (a) List all possible *shapes* of 2×3 matrices in RREF. Use $*$ to denote arbitrary entries. For example,

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

is one such shape. [**Hint:** There are 7 shapes in total.]

(b) List all possible shapes of 3×2 matrices in RREF.

(c) List all possible shapes of 3×3 matrices in RREF.

2.2.7. (a) Carry the matrix $\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$ to RREF without explicitly performing any row swaps, that is, without performing the elementary row operation $R_1 \leftrightarrow R_2$.

(b) State a sequence of elementary row operations that when performed will interchange the ith and jth rows of a matrix without explicitly swapping any rows.

(c) Verify that your sequence of elementary row operations given in part (b) is correct by showing that

$$\begin{bmatrix}\vdots\\R_i\\\vdots\\R_j\\\vdots\end{bmatrix} \longrightarrow \begin{bmatrix}\vdots\\R_j\\\vdots\\R_i\\\vdots\end{bmatrix}$$

under your sequence of row operations.

2.3 Rank

After solving numerous systems of equations, we are beginning to see the importance of leading entries in an REF of the augmented matrix of the system. This motivates the following definition.

Definition 2.3.1
Rank

The **rank** of a matrix A, denoted by $\text{rank}(A)$, is the number of leading entries in any REF of A.

If $\left[A \mid \vec{b}\right]$ is an augmented matrix, then $\text{rank}\left(\left[A \mid \vec{b}\right]\right)$ is the number of leading entries in any REF of $\left[A \mid \vec{b}\right]$.

Note that although we don't prove it here, given a matrix and any two of its REFs, the number of leading entries in both of these REFs will be the same. This means that our definition of rank actually makes sense.

Example 2.3.2

Consider the following three matrices A, B and C along with one of their REFs. Note that A and B are being viewed as augmented matrices for a linear system of equations, while C is being viewed as a coefficient matrix.

$$A = \left[\begin{array}{ccc|c} 2 & 1 & 9 & 31 \\ 0 & 1 & 2 & 8 \\ 1 & 0 & 3 & 10 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 3 & 10 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

$$B = \left[\begin{array}{cccc|c} 2 & 0 & 1 & 3 & 4 \\ 5 & 1 & 6 & -7 & 3 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 1 & 4 & -13 & -5 \\ 0 & -2 & -7 & 29 & 14 \end{array}\right]$$

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

We see that $\text{rank}(A) = 3$, $\text{rank}(B) = 2$ and $\text{rank}(C) = 1$.

Note that the requirement that a matrix be in REF before counting leading entries is important. The matrix

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$$

has two leading entries, but $\text{rank}(C) = 1$.

Exercise 25

Determine the ranks of the following matrices.

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 2 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that if a matrix A has m rows and n columns, then $\text{rank}(A) \leq \min\{m, n\}$, the minimum of m and n. This follows from the definition of leading entries and REF: there can be at most one leading entry in each row and each column.

The next theorem is useful to analyze systems of equations and will be used later in the course.

Theorem 2.3.3 **(System–Rank Theorem)**

Let $\left[A \mid \vec{b}\right]$ be the augmented matrix of a system of m linear equations in n variables.

(a) The system is consistent if and only if $\operatorname{rank}(A) = \operatorname{rank}\left(\left[A \mid \vec{b}\right]\right)$

(b) If the system is consistent, then the number of parameters in the general solution is the number of variables minus the rank of A:

$$\#\text{ of parameters} = n - \operatorname{rank}(A).$$

(c) The system is consistent for all $\vec{b} \in \mathbb{R}^m$ if and only if $\operatorname{rank}(A) = m$.

We don't prove the System–Rank Theorem here. However, we will look at some of the systems we have encountered thus far and show that they each satisfy all three parts of this Theorem.

Example 2.3.4 From Example 2.2.4, the system of $m = 3$ linear equations in $n = 3$ variables

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & 9x_3 & = & 31 \\ & & x_2 & + & 2x_3 & = & 8 \\ x_1 & & & + & 3x_3 & = & 10 \end{array}$$

has augmented matrix

$$\left[A \mid \vec{b}\right] = \left[\begin{array}{ccc|c} 2 & 1 & 9 & 31 \\ 0 & 1 & 2 & 8 \\ 1 & 0 & 3 & 10 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

and solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

From the System–Rank Theorem we see that

(a) $\operatorname{rank}(A) = 3 = \operatorname{rank}\left(\left[A \mid \vec{b}\right]\right)$ so the system is consistent.

(b) $\#$ of parameters $= n - \operatorname{rank}(A) = 3 - 3 = 0$ so there are no parameters in the solution (unique solution).

(c) $\operatorname{rank}(A) = 3 = m$ so the system will be consistent for any $\vec{b} \in \mathbb{R}^3$, that is, the system

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & 9x_3 & = & b_1 \\ & & x_2 & + & 2x_3 & = & b_2 \\ x_1 & & & + & 3x_3 & = & b_3 \end{array}$$

will be consistent (with a unique solution) for any choice of $b_1, b_2, b_3 \in \mathbb{R}$.

Example 2.3.5 From Example 2.2.8, the system of $m = 3$ linear equations in $n = 3$ variables

$$\begin{array}{rcrcrcr} 3x_1 & + & x_2 & & & = & 10 \\ 2x_1 & + & x_2 & + & x_3 & = & 6 \\ -3x_1 & + & 4x_2 & + & 15x_3 & = & -20 \end{array}$$

has augmented matrix

$$\left[\, A \mid \vec{b} \,\right] = \left[\begin{array}{rrr|r} 3 & 1 & 0 & 10 \\ 2 & 1 & 1 & 6 \\ -3 & 4 & 15 & -20 \end{array}\right] \longrightarrow \left[\begin{array}{rrr|r} 1 & 0 & -1 & 4 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

and solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

From the System–Rank Theorem, we have

(a) $\operatorname{rank}(A) = 2 = \operatorname{rank}\left(\left[\, A \mid \vec{b} \,\right]\right)$, so the system is consistent.

(b) # of parameters $= n - \operatorname{rank}(A) = 3 - 2 = 1$ so there is 1 parameter in the solution (infinitely many solutions).

(c) $\operatorname{rank}(A) = 2 \neq 3 = m$, so the system will not be consistent for every $\vec{b} \in \mathbb{R}^3$, that is, the system

$$\begin{array}{rcrcrcr} 3x_1 & + & x_2 & & & = & b_1 \\ 2x_1 & + & x_2 & + & x_3 & = & b_2 \\ -3x_1 & + & 4x_2 & + & 15x_3 & = & b_3 \end{array}$$

will be inconsistent for some choice of $b_1, b_2, b_3 \in \mathbb{R}$.

Example 2.3.6 From Example 2.2.10, the system of $m = 2$ equations in $n = 4$ variables

$$\begin{array}{rcrcrcrcr} x_1 & + & 6x_2 & & & - & x_4 & = & -1 \\ & & & & x_3 & + & 2x_4 & = & 7 \end{array}$$

has augmented matrix that is already in RREF

$$\left[\, A \mid \vec{b} \,\right] = \left[\begin{array}{rrrr|r} 1 & 6 & 0 & -1 & -1 \\ 0 & 0 & 1 & 2 & 7 \end{array}\right]$$

and solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 7 \\ 0 \end{bmatrix} + s \begin{bmatrix} -6 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}$$

From the System–Rank Theorem,

(a) $\operatorname{rank}(A) = 2 = \operatorname{rank}\left(\left[\, A \mid \vec{b} \,\right]\right)$, so the system is consistent.

(b) # of parameters $= n - \operatorname{rank}(A) = 4 - 2 = 2$ so there are 2 parameters in the solution (infinitely many solutions).

(c) $\operatorname{rank}(A) = 2 = m$, so the system will be consistent for every $\vec{b} \in \mathbb{R}^2$, that is, the system

$$\begin{array}{rcrcrcrcl} x_1 & + & 6x_2 & & & - & x_4 & = & b_1 \\ & & & & x_3 & + & 2x_4 & = & b_2 \end{array}$$

will be consistent (with infinitely many solutions) for any choice of $b_1, b_2 \in \mathbb{R}$.

Example 2.3.7 From Example 2.2.11, the system of $m = 3$ linear equations in $n = 3$ variables

$$\begin{array}{rcrcrcr} 2x_1 & + & 12x_2 & - & 8x_3 & = & -4 \\ 2x_1 & + & 13x_2 & - & 6x_3 & = & -5 \\ -2x_1 & - & 14x_2 & + & 4x_3 & = & 7 \end{array}$$

has augmented matrix

$$\left[\, A \mid \vec{b} \,\right] = \left[\begin{array}{rrr|r} 2 & 12 & -8 & -4 \\ 2 & 13 & -6 & -5 \\ -2 & -14 & 4 & 7 \end{array}\right] \longrightarrow \left[\begin{array}{rrr|r} 2 & 12 & -8 & -4 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

and is inconsistent. From the System–Rank Theorem, we see

(a) $\operatorname{rank}(A) = 2 < 3 = \operatorname{rank}\left(\left[\, A \mid \vec{b} \,\right]\right)$, so the system is inconsistent.

(b) as the system is inconsistent, the System–Rank Theorem does not apply here.

(c) $\operatorname{rank}(A) = 2 < 3 = m$ so the system will not be consistent for every $\vec{b} \in \mathbb{R}^3$. Indeed, as our work shows, the system is clearly not consistent for $\vec{b} = \left[\begin{smallmatrix} -4 \\ -5 \\ 7 \end{smallmatrix}\right]$.

In our last example, it is tempting to think that the system $\left[\, A \mid \vec{b} \,\right]$ will be inconsistent for every $\vec{b} \in \mathbb{R}^3$, however, this is not the case. If we take $\vec{b} = \vec{0}$, then our system becomes

$$\begin{array}{rcrcrcr} 2x_1 & + & 12x_2 & - & 8x_3 & = & 0 \\ 2x_1 & + & 13x_2 & - & 6x_3 & = & 0 \\ -2x_1 & - & 14x_2 & + & 4x_3 & = & 0 \end{array}$$

It isn't difficult to see that $x_1 = x_2 = x_3 = 0$ is a solution, so that this system is indeed consistent. Of course, we could ask for which $\vec{b} \in \mathbb{R}^3$ is this system consistent.

Example 2.3.8 Find an equation that $b_1, b_2, b_3 \in \mathbb{R}$ must satisfy so that the system

$$\begin{array}{rcrcrcl} 2x_1 & + & 12x_2 & - & 8x_3 & = & b_1 \\ 2x_1 & + & 13x_2 & - & 6x_3 & = & b_2 \\ -2x_1 & - & 14x_2 & + & 4x_3 & = & b_3 \end{array}$$

is consistent.

Solution: We carry the augmented matrix of this system to REF.

$$\left[\begin{array}{ccc|c} 2 & 12 & -8 & b_1 \\ 2 & 13 & -6 & b_2 \\ -2 & -14 & 4 & b_3 \end{array}\right] \underset{\substack{R_2-R_1 \\ R_3+R_1}}{\longrightarrow} \left[\begin{array}{ccc|c} 2 & 12 & -8 & b_1 \\ 0 & 1 & 2 & b_2-b_1 \\ 0 & -2 & -4 & b_3+b_1 \end{array}\right] \underset{R_3+2R_2}{\longrightarrow} \left[\begin{array}{ccc|c} 2 & 12 & -8 & b_1 \\ 0 & 1 & 2 & b_2-b_1 \\ 0 & 0 & 0 & -b_1+2b_2+b_3 \end{array}\right]$$

We see $\operatorname{rank}(A) = 2$ so we require $\operatorname{rank}\left(\left[A \mid \vec{b}\right]\right) = 2$ for consistency. Thus, we have

$$-b_1 + 2b_2 + b_3 = 0.$$

Note that if $-b_1 + 2b_2 + b_3 \neq 0$, then the above system is inconsistent.

It's possible that a linear system of equations may have coefficients which are defined in terms of a parameter (which we assume to be real numbers). Different values of these parameters will lead to different systems of linear equations. We can use the System–Rank Theorem to determine which values of the parameters will lead to systems with no solutions, one solution, and infinitely many solutions.

Example 2.3.9 For which values of the parameters $k, \ell \in \mathbb{R}$ does the system

$$\begin{array}{rcrcl} 2x_1 & + & 6x_2 & = & 5 \\ 4x_1 & + & (k+15)x_2 & = & \ell+8 \end{array}$$

have no solutions? A unique solution? Infinitely many solutions?

Solution: Let

$$A = \begin{bmatrix} 2 & 6 \\ 4 & k+15 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 5 \\ \ell+8 \end{bmatrix}.$$

We carry $\left[A \mid \vec{b}\right]$ to REF.

$$\left[\begin{array}{cc|c} 2 & 6 & 5 \\ 4 & k+15 & \ell+8 \end{array}\right] \underset{R_2-2R_1}{\longrightarrow} \left[\begin{array}{cc|c} 2 & 6 & 5 \\ 0 & k+3 & \ell-2 \end{array}\right]$$

We consider the cases $k+3 \neq 0$ and $k+3 = 0$.

- If $k+3 \neq 0$, that is if $k \neq -3$, then $\operatorname{rank}(A) = 2 = \operatorname{rank}\left(\left[A \mid \vec{b}\right]\right)$, so the system is consistent with $2 - \operatorname{rank}(A) = 2 - 2 = 0$ parameters. Hence we obtain a unique solution.
- If $k+3 = 0$, that is if $k = -3$, then

  $$\left[\begin{array}{cc|c} 2 & 6 & 5 \\ 0 & k+3 & \ell-2 \end{array}\right] \text{ simplifies as } \left[\begin{array}{cc|c} 2 & 6 & 5 \\ 0 & 0 & \ell-2 \end{array}\right].$$

  For $k = -3$, we must examine the cases $\ell - 2 \neq 0$ and $\ell - 2 = 0$

  - If $\ell - 2 \neq 0$, that is if $\ell \neq 2$, then $\operatorname{rank}(A) = 1 < 2 = \operatorname{rank}\left(\left[A \mid \vec{b}\right]\right)$ so the system is inconsistent and thus has no solutions.

- If $\ell - 2 = 0$, that is if $\ell = 2$, then $\text{rank}(A) = 1 = \text{rank}\left(\left[A \mid \vec{b}\right]\right)$ so the system is consistent with $2 - \text{rank}(A) = 2 - 1 = 1$ parameter. Hence we have infinitely many solutions.

In summary,

$$\begin{array}{rcl} \text{Unique Solution} & : & k \neq -3 \\ \text{No Solutions} & : & k = -3 \text{ and } \ell \neq 2 \\ \text{Infinitely Many Solutions} & : & k = -3 \text{ and } \ell = 2 \end{array}.$$

Definition 2.3.10

Underdetermined System of Linear Equations

A system of m linear equations in n variables is **underdetermined** if $n > m$, this is, if it has more variables than equations.

Example 2.3.11

The linear system of equations

$$\begin{array}{rcrcrcrcrl} x_1 & + & x_2 & - & x_3 & + & x_4 & - & x_5 & = 1 \\ x_1 & - & x_2 & - & 3x_3 & + & 2x_4 & + & 2x_5 & = 7 \end{array}$$

is underdetermined.

Theorem 2.3.12

A consistent underdetermined system of linear equations has infinitely many solutions.

Proof: Consider a consistent underdetermined system of m linear equations in n variables with coefficient matrix A. Since $\text{rank}(A) \leq \min\{m, n\} = m < n$, the system will have $n - \text{rank}(A) > 0$ parameters by part (b) of the System–Rank Theorem, and so will have infinitely many solutions. □

Definition 2.3.13

Overdetermined System of Linear Equations

A system of m linear equations in n variables is **overdetermined** if $n < m$, this is, if it has more equations than variables.

Example 2.3.14

The system of linear equations

$$\begin{array}{rcrcl} -2x_1 & + & x_2 & = & 2 \\ x_1 & - & 3x_2 & = & 4 \\ 3x_1 & + & 2x_2 & = & 7 \end{array}$$

is overdetermined.

Note that overdetermined linear systems are often inconsistent. Indeed, the system in the previous example is inconsistent. To see why this is, consider for example, three lines in $\mathbb{R}^2$ (so a system of three equations in two variables like the one in the previous example). When chosen arbitrarily, it is highly unlikely that all three lines would intersect in a common point and hence we would generally expect no solutions.

Section 2.3 Problems

2.3.1. Prove or disprove the following statements.

(a) A system of 3 equations in 5 variables has infinitely many solutions.

(b) A system of 5 equations in 3 variables cannot have infinitely many solutions.

(c) If the solution set to a system of equations is a line, then the coefficient matrix of the system has rank equal to 1.

(d) Let A be a matrix with m rows and n columns. If $\operatorname{rank}(A) < n$, then A has a row of zeros.

(e) Let A be a matrix with m rows and n columns. If $\operatorname{rank}(A) < m$, then A has a row of zeros.

2.3.2. (a) Give an example of a matrix with two rows and three columns whose rank is 1.

(b) Give an example of a matrix with three rows and two columns whose rank is 2.

(c) Is there an example of a matrix with four rows and two columns whose rank is 3? Either give such an example or explain why there cannot be one.

2.3.3. For which values of $\ell \in \mathbb{R}$ does the system

$$\begin{array}{rcrcrcr} 3x & - & 2y & + & 3z & = & 4 \\ 3x & + & 3y & + & 2z & = & 1 \\ -9x & - & 4y & + & (\ell^2 - 11)z & = & \ell - 4 \end{array}$$

have no solution? Exactly one solution? Infinitely many solutions? Justify your work.

2.3.4. Consider the system of equations

$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & + & 4x_3 & + & 8x_4 & = & 16 \\ x_1 & - & x_2 & & & + & x_4 & = & -1 \\ 3x_1 & + & & & 4x_3 & + & 10x_4 & = & k \end{array} \tag{2.1}$$

where $k \in \mathbb{R}$.

(a) Use the System–Rank Theorem to find the values of k such that (2.1) is consistent.

(b) For the value(s) of k found in part (a), use the System–Rank Theorem to determine the number of parameters in the solution to (2.1).

(c) For the value(s) of k found in part (a), find the solution to (2.1). Give the vector equation of the solution.

2.3.5. (a) Prove that if $ad - bc \neq 0$, then the reduced row echelon form of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
[**Hint:** Consider the cases $a = 0$ and $a \neq 0$ separately.]

(b) Deduce that if $ad - bc \neq 0$, then the linear system

$$\begin{array}{rcrcr} ax & + & by & = & p \\ cx & + & dy & = & q \end{array}$$

has a unique solution.

2.3.6. Consider a parabola in $\mathbb{R}^2$ given by the equation $y = a_0 + a_1 x + a_2 x^2$ where $a_0, a_1, a_2 \in \mathbb{R}$ with $a_2 \neq 0$. Let $P(0,0)$, $Q(1,0)$ and $R(c,d)$ be points in $\mathbb{R}^2$ where $c, d \in \mathbb{R}$ are arbitrary. Using the System–Rank Theorem, determine all c and d so that

(a) there is exactly one such parabola passing through the points P, Q and R.

(b) there are infinitely many such parabolas passing through the points P, Q and R.

(c) there is no such parabola passing through the points P, Q and R.

2.3.7. Consider two planes T_1 and T_2 in $\mathbb{R}^3$ given by the scalar equations

$$n_1 x_1 + n_2 x_2 + n_3 x_3 = a_1 \quad \text{and} \quad m_1 x_1 + m_2 x_2 + m_3 x_3 = a_2,$$

respectively. Using the System–Rank Theorem, show that these two planes either never intersect, or that they intersect in a line or a plane. [**Hint**: Consider two cases – one with T_1 and T_2 parallel, and one with T_1 and T_2 not parallel.]

2.4 Homogeneous Systems of Linear Equations

We now discuss a particular type of linear system of equations that have some very nice properties.

Definition 2.4.1

Homogeneous System of Linear Equations

A **homogeneous linear equation** is a linear equation where the constant term is zero. A **system of homogeneous linear equations** is a collection of finitely many homogeneous linear equations.

A homogeneous system of m linear equations in n variables is written as

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

As this is still a linear system of equations, we use our usual techniques to solve such systems. However, notice that $x_1 = x_2 = \cdots = x_n = 0$ satisfies each equation in the homogeneous system, and thus $\vec{0} \in \mathbb{R}^n$ is a solution to this system, called the *trivial solution*. As every homogeneous system has a trivial solution, we see immediately that homogeneous linear systems of equations are always consistent.

Example 2.4.2

Solve the homogeneous linear system

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 0 \\ & & 3x_2 & - & x_3 & = & 0 \end{array}.$$

Solution: We have

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 3 & -1 & 0 \end{array}\right] \underset{\frac{1}{3}R_2}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -1/3 & 0 \end{array}\right] \overset{R_1-R_2}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 0 & 4/3 & 0 \\ 0 & 1 & -1/3 & 0 \end{array}\right]$$

so

$$\begin{array}{ccc} x_1 & = & -\frac{4}{3}t \\ x_2 & = & \frac{1}{3}t \\ x_3 & = & t \end{array}, \quad t \in \mathbb{R} \qquad \text{or} \qquad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -4/3 \\ 1/3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

We make a few remarks about Example 2.4.2:

- Note that taking $t = 0$ gives the trivial solution, which is just one of infinitely many solutions for the system. This should not be surprising since our system is underdetermined and consistent (consistency follows from the system being homogeneous). Indeed, the solution set is actually a line through the origin.
- We can simplify our solution a little bit by eliminating fractions:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -4/3 \\ 1/3 \\ 1 \end{bmatrix} = \frac{t}{3} \begin{bmatrix} -4 \\ 1 \\ 3 \end{bmatrix} = s \begin{bmatrix} -4 \\ 1 \\ 3 \end{bmatrix}, \quad s \in \mathbb{R}$$

  where $s = t/3$. Hence we can let the parameter "absorb" the factor of $1/3$. This is not necessary, but is useful if one wishes to eliminate fractions.

- When row reducing the augmented matrix of a homogeneous systems of linear equations, notice that the last column always contains zeros regardless of the row operations performed. Thus, it is common to row reduce only the coefficient matrix:

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & -1 \end{bmatrix} \underset{\frac{1}{3}R_2}{\longrightarrow} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1/3 \end{bmatrix} \underset{\longrightarrow}{R_1-R_2} \begin{bmatrix} 1 & 0 & 4/3 \\ 0 & 1 & -1/3 \end{bmatrix}.$$

Definition 2.4.3

Associated Homogeneous System of Linear Equations

Given a non-homogeneous linear system of equations with augmented matrix $\left[A \mid \vec{b}\right]$ (so $\vec{b} \neq \vec{0}$), the homogeneous system with augmented matrix $\left[A \mid \vec{0}\right]$ is called the **associated homogeneous system**.

The solution to the associated homogeneous system tells us a lot about the solution of the original non-homogeneous system. If we solve the system

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 1 \\ & & 3x_2 & - & x_3 & = & 3 \end{array} \tag{2.2}$$

we have

$$\left[\begin{array}{rrr|r} 1 & 1 & 1 & 1 \\ 0 & 3 & -1 & 3 \end{array}\right] \underset{\frac{1}{3}R_2}{\longrightarrow} \left[\begin{array}{rrr|r} 1 & 1 & 1 & 1 \\ 0 & 1 & -1/3 & 1 \end{array}\right] \underset{\longrightarrow}{R_1-R_2} \left[\begin{array}{rrr|r} 1 & 0 & 4/3 & 0 \\ 0 & 1 & -1/3 & 1 \end{array}\right]$$

so

$$\begin{array}{rcr} x_1 & = & -\frac{4}{3}t \\ x_2 & = & 1+\frac{1}{3}t, \\ x_3 & = & t \end{array} \quad t \in \mathbb{R} \qquad \text{or} \qquad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -4/3 \\ 1/3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

Recall that the solution to the associated homogeneous system from Example 2.4.2 is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -4/3 \\ 1/3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

so we view the homogeneous solution from Example 2.4.2 as a line, say L_0, through the origin, and the solution from (2.2) as a line, say L_1, through $P(0,1,0)$ parallel to L_0. We refer to $\left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right]$ as a *particular solution* to (2.2) and note that in general, the solution to a consistent non-homogeneous system of linear equations is a particular solution to that system plus the general solution to the associated homogeneous system of linear equations.

$$\overbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\substack{\text{particular} \\ \text{solution}}} + \underbrace{t \begin{bmatrix} -4/3 \\ 1/3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}}_{\substack{\text{associated} \\ \text{homogeneous} \\ \text{solution}}}}^{\text{solution to the system of equations}} \qquad \overbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -4/3 \\ 1/3 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}}^{\substack{\text{solution to the} \\ \text{associated homogeneous system of equations}}}$$

What we have observed here is true for any system of linear equations. We state this result as a theorem, but we omit the proof.

Theorem 2.4.4

Let $\vec{x}_0$ be a particular solution to a given system of linear equations. Then $\vec{x}_0 + \vec{s}$ is a solution to this system if and only if $\vec{s}$ is a solution to the associated homogeneous system of linear equations.

Example 2.4.5 Consider the system of linear equations

$$\begin{array}{rcrcrcrcr} x_1 & + & 6x_2 & & & - & x_4 & = & -1 \\ & & & & x_3 & + & 2x_4 & = & 7 \end{array}$$

We know from Example 2.2.10 that the solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 7 \\ 0 \end{bmatrix} + s \begin{bmatrix} -6 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R},$$

which is as a plane through $(-1, 0, 7, 0)$ in $\mathbb{R}^4$ since the vectors $\begin{bmatrix} -6 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}$ are nonzero and not parallel. Thus the solution to the associated homogeneous system

$$\begin{array}{rcrcrcrcr} x_1 & + & 6x_2 & & & - & x_4 & = & 0 \\ & & & & x_3 & + & 2x_4 & = & 0 \end{array}$$

is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = s \begin{bmatrix} -6 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}.$$

which we recognize as a plane through the origin in $\mathbb{R}^4$.

Another nice property of homogeneous systems of linear equations is that given two solutions, say $\vec{x}_1$ and $\vec{x}_2$, any linear combination of them is also a solution to the system.

Example 2.4.6 Consider a homogeneous system of m linear equations in n unknowns. Suppose $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ and $\vec{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$ are solutions to this system. Show that $c_1\vec{y} + c_2\vec{z}$ is also a solution to this system for any $c_1, c_2 \in \mathbb{R}$.

Proof: Since $\vec{y}$ and $\vec{z}$ satisfy the homogeneous system of linear equations, they satisfy any arbitrary equation of the system, say $a_1x_1 + \cdots + a_nx_n = 0$. Thus we have that

$$a_1y_1 + \cdots + a_ny_n = 0 = a_1z_1 + \cdots + a_nz_n.$$

We verify that

$$c_1\vec{y} + c_2\vec{z} = \begin{bmatrix} c_1y_1 + c_2z_1 \\ \vdots \\ c_1y_n + c_2z_n \end{bmatrix}$$

satisfies this arbitrary equation as well. We have

$$\begin{aligned} a_1(c_1y_1 + c_2z_1) + \cdots + a_n(c_1y_n + c_2z_n) &= c_1(a_1y_1 + \cdots + a_ny_n) + c_2(a_1z_1 + \cdots + a_nz_n) \\ &= c_1(0) + c_2(0) \\ &= 0. \end{aligned}$$

Hence $c_1\vec{y} + c_2\vec{z}$ is also a solution to the homogeneous system of linear equations.

Section 2.4 Problems

2.4.1. Find the solutions to the following homogeneous systems of linear equations. Compare your answers to the solutions you found in Problem 2.2.1.

(a) $\begin{array}{rcrcl} 6x_1 & + & 7x_2 & = & 0 \\ 3x_1 & + & 2x_2 & = & 0 \end{array}$.

(b) $\begin{array}{rcrcrcrcl} 2x_1 & - & 2x_2 & + & x_3 & + & 15x_4 & = & 0 \\ 3x_1 & - & 3x_2 & + & x_3 & + & 21x_4 & = & 0 \end{array}$.

(c) $\begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & 3x_3 & = & 0 \\ x_1 & + & 3x_2 & + & x_3 & = & 0 \\ 2x_1 & + & 2x_2 & + & 10x_3 & = & 0 \end{array}$.

(d) $\begin{array}{rcrcrcl} x_1 & + & 3x_2 & + & 10x_3 & = & 0 \\ 2x_1 & + & 6x_2 & + & 19x_3 & = & 0 \\ -x_1 & - & 2x_2 & - & 6x_3 & = & 0 \end{array}$.

(e) $\begin{array}{rcrcrcrcrcl} 2x_1 & - & 4x_2 & + & x_3 & + & 2x_4 & + & 3x_5 & = & 0 \\ x_1 & - & 2x_2 & + & x_3 & + & 2x_4 & + & x_5 & = & 0 \\ 3x_1 & - & 6x_2 & + & 3x_3 & + & 3x_4 & + & 6x_5 & = & 0 \end{array}$.

2.4.2. Prove or disprove the following statements.

(a) If a system of linear equations has $\vec{s} = \vec{0}$ as a solution, then the system must be homogeneous.

(b) If a system of linear equations has $\vec{s} \neq \vec{0}$ as a solution, then the system cannot be homogeneous.

2.4.3. Suppose that $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ are solutions to some homogeneous system. Explain why $\begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$ must be a solution to that same homogeneous system.

2.4.4. Let $A = \begin{bmatrix} 1 & 4 & 3 \\ 2 & 8 & 3 \\ -3 & -12 & 4 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 8 \\ 13 \\ -11 \end{bmatrix}$.

(a) Determine the solution set of the homogeneous system with coefficient matrix A.

(b) Show that $\vec{s} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is a solution to the non-homogeneous system $\left[A \mid \vec{b} \right]$.

(c) Use the results in parts (a) and (b) to find the solution set of the non-homogeneous system $\left[A \mid \vec{b} \right]$.

2.5 Comments on Combining Elementary Row Operations

Having performed many elementary row operations by this point, it's a good idea to review some rules about combining elementary row operations, that is, performing multiple elementary row operations in the same step. Many of the previous examples contain instances where systems are solved by performing multiple row operations to the augmented matrix in the same step. For example,

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 2 & 1 & 1 & 6 \\ -3 & 4 & 15 & -20 \end{array}\right] \xrightarrow[\substack{R_2-2R_1 \\ R_3+3R_1}]{} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 0 & 1 & 3 & -2 \\ 0 & 4 & 12 & -8 \end{array}\right].$$

Here we are simply using one row to modify the other rows. This is completely acceptable (and encouraged) since we only have to write out matrices twice as opposed to three times. We must be careful however, as not all elementary row operations can be combined. Consider the following linear system of equations.

$$\begin{array}{ccccc} x_1 & + & x_2 & = & 1 \\ x_1 & - & x_2 & = & -1 \end{array}.$$

If we perform the following operations

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & -1 & -1 \end{array}\right] \xrightarrow{\substack{R_1-R_2 \\ R_2-R_1}} \left[\begin{array}{cc|c} 0 & 2 & 2 \\ 0 & -2 & -2 \end{array}\right] \xrightarrow[R_2+R_1]{} \left[\begin{array}{cc|c} 0 & 2 & 2 \\ 0 & 0 & 0 \end{array}\right] \xrightarrow{\frac{1}{2}R_1} \left[\begin{array}{cc|c} 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right],$$

then we find that

$$\vec{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}$$

appears to be the solution. However, this is incorrect since the system has the unique solution $\vec{x} = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$. The error occurs in the first set of row operations. Here both the first and second rows are used to modify the other. If we perform $R_1 - R_2$ to R_1, then we have now changed the first row. If we then go on to perform $R_2 - R_1$ to R_2, then we should use the updated R_1 and not the original R_1. Thus we should separate our first step above into two steps:

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & -1 & -1 \end{array}\right] \xrightarrow{R_1-R_2} \left[\begin{array}{cc|c} 0 & 2 & 2 \\ 1 & -1 & -1 \end{array}\right] \xrightarrow[R_2-R_1]{} \left[\begin{array}{cc|c} 0 & 2 & 2 \\ 1 & -3 & -3 \end{array}\right] \longrightarrow \cdots.$$

Clearly, this is not the best choice of row operations to solve the system! However the goal of this example is not to find a solution, but rather illustrate that we should not modify a given row in one step while at the same time using it to modify another row.

Another thing to avoid is modifying a row multiple times in the same step. This itself is not mathematically wrong, but is generally shunned as it often leads students to arithmetic errors. For example, while

$$\begin{bmatrix} 2 & 1 & 3 \\ 6 & 2 & 4 \\ 18 & 5 & 7 \end{bmatrix} \xrightarrow[R_3+3R_1-4R_2]{} \begin{bmatrix} 2 & 1 & 3 \\ 6 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

is mathematically correct, it is not immediately obvious that such a row operation would be useful, and it forces the student to do more "mental math" which often leads to mistakes. A better option would be

$$\begin{bmatrix} 2 & 1 & 3 \\ 6 & 2 & 4 \\ 18 & 5 & 7 \end{bmatrix} \xrightarrow[\substack{R_2-3R_1 \\ R_3-9R_1}]{} \begin{bmatrix} 2 & 1 & 3 \\ 0 & -1 & -5 \\ 0 & -4 & -20 \end{bmatrix} \xrightarrow[R_3-4R_2]{} \begin{bmatrix} 2 & 1 & 3 \\ 0 & -1 & -5 \\ 0 & 0 & 0 \end{bmatrix}$$

which is more natural and has simpler computations.

To summarize, students are encouraged to combine row operations as it leads to less writing and shorter solutions. However, keep in mind that on any given step, one must not modify a given row while using that row to modify another row, and that one should avoid modifying a row more than once in the same step.

Chapter 3

Matrices

3.1 Matrix Algebra

We first encountered matrices when we solved systems of linear equations, where we performed elementary row operations to the augmented matrix or the coefficient matrix of the system. We now consider matrices as algebraic objects, defining the operations of matrix addition and scalar multiplication. Our work here will closely mirror that of Section 1.1, and we will see that under these operations, matrices behave much like vectors in $\mathbb{R}^n$.

Definition 3.1.1 **Matrix, (i,j)-entry, $M_{m\times n}(\mathbb{R})$, Square Matrix**

An $m \times n$ **matrix** A is a rectangular array with m rows and n columns. The entry in the ith row and jth column will be denoted by either a_{ij} or $(A)_{ij}$, that is

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

We sometimes abbreviate this as $A = [a_{ij}]$.

The set of all $m \times n$ matrices with real entries is denoted by $M_{m\times n}(\mathbb{R})$. For a matrix $A \in M_{m\times n}(\mathbb{R})$, we say that A has **size** $m \times n$ and call a_{ij} the (i,j)**-entry** of A. If $m = n$, we say that A is a **square matrix**.

Note that the rows of a matrix are labeled from top to bottom, and the columns are labeled from left to right.

Example 3.1.2

Let

$$A = \begin{bmatrix} 1 & 2 \\ 6 & 4 \\ 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 7 & 3 \\ -2 & 3 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 5 & 0 \\ 0 & \sqrt{2} \end{bmatrix}.$$

Then A is a 3×2 matrix, B is a 2×3 matrix, and C is a *square* 2×2 matrix. That is, $A \in M_{3\times 2}(\mathbb{R})$, $B \in M_{2\times 3}(\mathbb{R})$ and $C \in M_{2\times 2}(\mathbb{R})$.

The $(1,2)$-entry of A is 2. The $(2,3)$-entry of B is 0.

Example 3.1.3

A matrix of size $m \times 1$ is of the form

$$A = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}.$$

We call this a *column matrix* (or *column vector*). We see therefore that $M_{m\times 1}(\mathbb{R}) = \mathbb{R}^m$.

A matrix of size $1 \times n$ is of the form

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \end{bmatrix}.$$

We call this a *row matrix* (or *row vector*).

Finally, a matrix of size 1×1 is of the form

$$C = \begin{bmatrix} c \end{bmatrix}.$$

Occasionally we will identify a 1×1 matrix with a real number (that is, we will view $C = \begin{bmatrix} c \end{bmatrix}$ as though it were simply $c \in \mathbb{R}$), even though technically these are different objects.

Definition 3.1.4
Zero Matrix

The $m \times n$ matrix with all zero entries is called a **zero matrix**, denoted by $0_{m\times n}$, or simply by 0 if the size is clear.

Example 3.1.5

We have

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

We will now introduce some basic algebraic operations that can be performed on matrices. We start by defining what it means for two matrices to be equal.

Definition 3.1.6
Matrix Equality

Two matrices $A = [a_{ij}] \in M_{m\times n}(\mathbb{R})$ and $B = [b_{ij}] \in M_{p\times k}(\mathbb{R})$ are **equal** if $m = p$, $n = k$ and $a_{ij} = b_{ij}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. We denote this by $A = B$. We write $A \neq B$ when A and B are not equal.

That is, two matrices are equal if and only if they have the same size and their corresponding entries are equal.

Example 3.1.7

If $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is equal to $B = \left[\begin{smallmatrix} 7 & -2 \\ 0 & 5 \end{smallmatrix}\right]$, then

$$a = 7, \quad b = -2, \quad c = 0 \quad \text{and} \quad d = 5.$$

Example 3.1.8

No two of the matrices

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

are equal, since they have different sizes.

Next, we define addition, subtraction and scalar multiplication of matrices.

Definition 3.1.9

Matrix Addition, Matrix Subtraction, Matrix Scalar Multiplication

Let $A, B \in M_{m\times n}(\mathbb{R})$.

Matrix addition is defined by letting $A + B$ be the $m \times n$ matrix whose (i, j)-entry is

$$(A + B)_{ij} = (A)_{ij} + (B)_{ij}.$$

Matrix subtraction is defined by letting $A - B$ be the $m \times n$ matrix whose (i, j)-entry is

$$(A - B)_{ij} = (A)_{ij} - (B)_{ij}.$$

For $c \in \mathbb{R}$, the **scalar multiple** cA is the $m \times n$ matrix whose (i, j)-entry is

$$(cA)_{ij} = c(A)_{ij}.$$

That is, the entries of $A+B$ are the sums of the corresponding entries of A and B, and the entries of $A - B$ are the differences of the corresponding entries of A and B. Likewise, the entries of cA are the the entries of A multiplied by c. It is important to keep in mind that matrix addition and subtraction are only defined for matrices of the same size. Also note that $A - B = A + (-1)B$.

Example 3.1.10

Let

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

Compute $A + B$, $A - B$ and $5A$.

Solution: We have

$$\begin{aligned} A + B &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 4 \\ 2 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1+0 & 2+1 & 3+4 \\ 0+2 & 1+3 & -1+3 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 7 \\ 2 & 4 & 2 \end{bmatrix}, \\ A - B &= \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 4 \\ 2 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1-0 & 2-1 & 3-4 \\ 0-2 & 1-3 & -1-3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ -2 & -2 & -4 \end{bmatrix}, \\ 5A &= 5\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 5(1) & 5(2) & 5(3) \\ 5(0) & 5(1) & 5(-1) \end{bmatrix} = \begin{bmatrix} 5 & 10 & 15 \\ 0 & 5 & -5 \end{bmatrix}. \end{aligned}$$

Example 3.1.11 The expressions

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 3 & 4 \\ 0 & 0 & 1 \\ 3 & 6 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 6 & 4 \end{bmatrix}$$

are undefined since the matrices involved have different sizes.

Exercise 26 Find $a, b, c \in \mathbb{R}$ such that

$$\begin{bmatrix} a & b & c \end{bmatrix} - 2\begin{bmatrix} c & a & b \end{bmatrix} = \begin{bmatrix} -3 & 3 & 6 \end{bmatrix}.$$

It follows from our definition of scalar multiplication that for any $A \in M_{m\times n}(\mathbb{R})$ and $c \in \mathbb{R}$

$$0A = 0_{m\times n} \quad \text{and} \quad c\,0_{m\times n} = 0_{m\times n}.$$

The next example shows that if $cA = 0_{m\times n}$, then either $c = 0$ or $A = 0_{m\times n}$.

Example 3.1.12 Let $c \in \mathbb{R}$ and $A \in M_{m\times n}(\mathbb{R})$ be such that $cA = 0_{m\times n}$. Prove that either $c = 0$ or $A = 0_{m\times n}$.

Proof: Since $cA = 0_{m\times n}$, we have that

$$ca_{ij} = 0 \text{ for every } i = 1, \ldots, m \text{ and } j = 1, \ldots, n. \tag{3.1}$$

If $c = 0$, then the result holds, so we assume $c \neq 0$. But then from (3.1), we see that $a_{ij} = 0$ for every $i = 1, \ldots, m$ and $j = 1, \ldots, n$, that is, $A = 0_{m\times n}$. $\square$

The next theorem is very similar to the Fundamental Properties of Vector Algebra, and shows that under our operations of addition and scalar multiplication, matrices behave much like vectors in $\mathbb{R}^n$.

Theorem 3.1.13 **(Fundamental Properties of Matrix Algebra)**

Let $A, B, C \in M_{m\times n}(\mathbb{R})$ and let $c, d \in \mathbb{R}$. We have

M1. $A + B \in M_{m\times n}(\mathbb{R})$ — $M_{m\times n}(\mathbb{R})$ is closed under addition

M2. $A + B = B + A$ — addition is commutative

M3. $(A + B) + C = A + (B + C)$ — addition is associative

M4. $cA \in M_{m\times n}(\mathbb{R})$ — $M_{m\times n}(\mathbb{R})$ is closed under scalar multiplication

M5. $c(dA) = (cd)A$ — scalar multiplication is associative

M6. $(c + d)A = cA + dA$ — distributive law

M7. $c(A + B) = cA + cB$ — distributive law

We close this section with another operation that we can perform on matrices. This operation will seem strange now, but we will learn later that it can be very useful.

Definition 3.1.14
Transpose of a Matrix

Let $A \in M_{m\times n}(\mathbb{R})$. The **transpose** of A, denoted by A^T, is the $n \times m$ matrix satisfying $(A^T)_{ij} = (A)_{ji}$.

That is, the rows of A^T are the columns of A.

Example 3.1.15

Let

$$A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 4 & 8 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 4 & 2 \\ -1 & 3 \end{bmatrix}.$$

Then

$$A^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \quad B^T = \begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix} \quad \text{and} \quad C^T = \begin{bmatrix} 4 & -1 \\ 2 & 3 \end{bmatrix}.$$

Theorem 3.1.16

(Properties of Transpose)

Let $A, B \in M_{m\times n}(\mathbb{R})$ and $c \in \mathbb{R}$. Then

(a) $A^T \in M_{n\times m}(\mathbb{R})$.

(b) $\left(A^T\right)^T = A$.

(c) $(A + B)^T = A^T + B^T$.

(d) $(cA)^T = cA^T$.

Exercise 27

Prove that $(A - B)^T = A^T - B^T$.

Example 3.1.17

Solve for $A \in M_{2\times 2}(\mathbb{R})$ if

$$\left(2A^T - 3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}\right)^T = \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix}.$$

Solution: Using Theorem 3.1.16, we have

$$\begin{aligned}
\left(2A^T\right)^T - \left(3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}\right)^T &= \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} && \text{by (c)} \\
2\left(A^T\right)^T - 3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}^T &= \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} && \text{by (d)} \\
2A - 3\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} &= \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} && \text{by (b)}
\end{aligned}$$

$$\begin{aligned}
2A &= \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} + \begin{bmatrix} 3 & -3 \\ 6 & 3 \end{bmatrix} \\
A &= \frac{1}{2}\begin{bmatrix} 5 & 0 \\ 5 & 5 \end{bmatrix} \\
A &= \begin{bmatrix} 5/2 & 0 \\ 5/2 & 5/2 \end{bmatrix}.
\end{aligned}$$

Exercise 28 Give examples of nonzero matrices $A, B \in M_{2\times 2}(\mathbb{R})$ such that

$$A^T = A \quad \text{and} \quad B^T = -B.$$

Section 3.1 Problems

3.1.1. Let

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 4 & 6 \\ 11 & -4 & 10 \\ 5 & 8 & -2 \end{bmatrix}.$$

Determine a_{11}, a_{23} and a_{42}.

3.1.2. Let

$$A = \begin{bmatrix} 2 & -3 & 3 & 2 \\ 4 & -3 & 4 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -2 & 1 & 1 \\ 2 & 1 & -3 & 5 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 4 & -2 & -3 & 1 \\ -5 & 1 & 2 & 2 \end{bmatrix}.$$

(a) Compute $4A - 3B$.

(b) Compute $(A + B + C)^T$.

3.1.3. Solve for the matrix A if

$$4A - 9\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} = \left(2A^T - 5\begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}\right)^T.$$

3.1.4. A square matrix $B \in M_{n\times n}(\mathbb{R})$ is said to be **symmetric** if $B^T = B$.

(a) Show that $A + A^T$ is symmetric for all $A \in M_{n\times n}(\mathbb{R})$.

(b) Give conditions on $s, t \in \mathbb{R}$ such that the matrix $C = \begin{bmatrix} s & t \\ st & 1 \end{bmatrix}$ is symmetric.

3.1.5. A square matrix $B \in M_{n\times n}(\mathbb{R})$ is said to be **skew-symmetric** if $B^T = -B$.

(a) Give an example of a *non-zero* 3×3 skew-symmetric matrix.

(b) Show that $A - A^T$ is skew-symmetric for all $A \in M_{n\times n}(\mathbb{R})$.

3.1.6. (a) Show that every $A \in M_{n\times n}(\mathbb{R})$ can be expressed as the sum of a symmetric matrix and a skew-symmetric matrix.
[**Hint:** Look at the previous two problems and consider $A \pm A^T$.]

(b) Let $A = \left[\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right]$. Express A as the sum of a symmetric matrix and a skew-symmetric matrix.

3.2 The Matrix–Vector Product

In this section, we define the product of a matrix and a vector and explore the algebraic properties of this product. In the next section, we will see how this product can be used to better understand properties of systems of linear equations.

In order to define the matrix–vector product, we need to describe the entries of a matrix in a slightly different way. Thus far, we have expressed matrices in terms of their explicit entries using $A = [a_{ij}]$, but this is not always necessary or desirable. In what follows, we will want to consider a matrix in terms of its columns. For example, consider the matrix

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

If we define

$$\vec{a}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} 3 \\ -4 \end{bmatrix} \quad \text{and} \quad \vec{a}_3 = \begin{bmatrix} -2 \\ 3 \end{bmatrix},$$

then we can express A more compactly as $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix}$.

More generally, if $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^m$, then $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$ is a matrix in $M_{m\times n}(\mathbb{R})$ whose jth column of A is $\vec{a}_j$. Specifically, the (i, j)-entry of A is the ith entry of $\vec{a}_j$.

Example 3.2.1 If

$$A = \begin{bmatrix} 1 & 0 & 5 \\ -1 & 0 & 4 \\ 3 & 2 & -3 \\ 4 & -3 & 3 \end{bmatrix},$$

then we may write $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix}$ where

$$\vec{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \\ 4 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \vec{a}_3 = \begin{bmatrix} 5 \\ 4 \\ -3 \\ 3 \end{bmatrix}.$$

Notice that $A \in M_{4\times 3}(\mathbb{R})$ so each column of A belongs to $\mathbb{R}^4$. The $(2, 3)$-entry of A is 4, which is the second entry of $\vec{a}_3$.

Definition 3.2.2
Matrix–Vector Product

Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{m\times n}(\mathbb{R})$ and $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$. Then the vector $A\vec{x}$ is defined by

$$A\vec{x} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n \in \mathbb{R}^m.$$

In words, given $A \in M_{m\times n}(\mathbb{R})$ and $\vec{x} \in \mathbb{R}^n$, the matrix–vector product of $A\vec{x}$ is simply a linear combination of the columns of A where the entries in the vector $\vec{x}$ are the coefficients or scalars in the linear combination. Since $A \in M_{m\times n}(\mathbb{R})$, the columns of A are vectors in $\mathbb{R}^m$ and thus $A\vec{x} \in \mathbb{R}^m$.

The order is important in the matrix–vector product: it is incorrect to write $\vec{x}A$. It's also important to keep the sizes of our matrices and vectors in mind. For $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{m\times n}(\mathbb{R})$, the matrix-vector product $A\vec{x}$ only makes sense when $\vec{x} \in \mathbb{R}^n$:

$$\underbrace{A}_{\mathbf{m}\times\mathbf{n}}\underbrace{\vec{x}}_{\mathbb{R}^{\mathbf{n}}} = \underbrace{x_1\vec{a}_1 + \cdots + x_n\vec{a}_n}_{\mathbb{R}^{\mathbf{m}}}$$

Thus, the number of components of $\vec{x}$ must be equal to the number of columns in A in order for the matrix–vector product $A\vec{x}$ to be defined. For example, the matrix–vector product

$$A\vec{x} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

is not defined since A has two columns but $\vec{x} \notin \mathbb{R}^2$.

Example 3.2.3

Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -2 & 1 & 0 \\ 1 & 0 & 5 & 2 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} -1 \\ 1 \\ 2 \\ -1 \end{bmatrix}.$$

Compute $A\vec{x}$ and $B\vec{y}$.

Solution: We have

$$A\vec{x} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ -3 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 3 \end{bmatrix} - 3\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} -4 \\ -6 \end{bmatrix}$$

and

$$B\vec{y} = \begin{bmatrix} 3 & -2 & 1 & 0 \\ 1 & 0 & 5 & 2 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \\ 2 \\ -1 \end{bmatrix} = (-1)\begin{bmatrix} 3 \\ 1 \end{bmatrix} + 1\begin{bmatrix} -2 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 5 \end{bmatrix} - 1\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ 7 \end{bmatrix}.$$

Exercise 29

Compute the following products.

(a) $\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

(b) $\begin{bmatrix} 2 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

(c) $A\vec{0}$, where $A \in M_{m\times n}(\mathbb{R})$ and $\vec{0} = \vec{0}_{\mathbb{R}^n}$.

Example 3.2.4 Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Compute $A\vec{e}_1$, $A\vec{e}_2$ and $A\vec{e}_3$.

Solution: We have

$$A\vec{e}_1 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = (1) \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} + (0) \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} + (0) \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix},$$

$$A\vec{e}_2 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = (0) \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} + (1) \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} + (0) \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix},$$

$$A\vec{e}_3 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = (0) \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} + (0) \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} + (1) \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}.$$

Notice that in Example 3.2.4 the product $A\vec{e}_i$ returned the ith column of A. In the next exercise you are asked to generalize this to the case of an arbitrary $m \times n$ matrix.

Exercise 30 Let $A \in M_{m\times n}(\mathbb{R})$, and let $\vec{e}_i$ be the vector in $\mathbb{R}^n$ whose ith component is 1 and whose remaining components are all 0. Show that $A\vec{e}_i = \vec{a}_i$, the ith column of A.

The next two examples highlight one feature of matrix–vector multiplication that is unlike real number multiplication.

Example 3.2.5 Let

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Then since

$$A\vec{x} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

we see that despite $A \in M_{m\times n}(\mathbb{R})$ and $\vec{x} \in \mathbb{R}^n$ both being nonzero, we have $A\vec{x} = \vec{0}$.

Example 3.2.5 will likely seem strange. For nonzero $a, x \in \mathbb{R}$, we know that $ax \neq 0$. As we continue to define new algebraic objects and then adapt our usual operations of addition and scalar multiplication to work with these new objects, we will need to be on the lookout for strange situations such as this[1].

[1] Recall the cross product in $\mathbb{R}^3$ also had some strange properties.

Example 3.2.6 Let

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -1 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Then

$$A\vec{x} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}$$

and

$$B\vec{x} = \begin{bmatrix} 3 & -1 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1\begin{bmatrix} 3 \\ 2 \end{bmatrix} + 2\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}.$$

We see that $A\vec{x} = B\vec{x}$ with $\vec{x} \neq \vec{0}$, and yet $A \neq B$.

Example 3.2.6 might again seem strange. For $a, b, x \in \mathbb{R}$ with $x \neq 0$, we know that if $ax = bx$, then $a = b$. As Example 3.2.6 shows, this result does not hold for the matrix–vector product: $A\vec{x} = B\vec{x}$ for a given nonzero vector $\vec{x}$ is *not* sufficient to guarantee $A = B$.

Theorem 3.2.7 **(Matrix Equality Theorem)**

Let $A, B \in M_{m\times n}(\mathbb{R})$. If $A\vec{x} = B\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$, then $A = B$.

Note that the hypothesis of the Matrix Equality Theorem requires $A\vec{x} = B\vec{x}$ *for every* $\vec{x} \in \mathbb{R}^n$. Our result from Example 3.2.6 does not contradict the Matrix Equality Theorem, as there we only had that $A\vec{x} = B\vec{x}$ *for some* $\vec{x} \in \mathbb{R}^2$, namely $\vec{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$.

Proof (of the Matrix Equality Theorem): Let $A, B \in M_{m\times n}(\mathbb{R})$ with

$$A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_n \end{bmatrix}.$$

Since $A\vec{x} = B\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$, we have that $A\vec{e}_i = B\vec{e}_i$ for $i = 1, \ldots, n$. Since

$$A\vec{e}_i = \vec{a}_i \quad \text{and} \quad B\vec{e}_i = \vec{b}_i$$

(see Exercise 30) we have that $\vec{a}_i = \vec{b}_i$ for $i = 1, \ldots, n$. Hence $A = B$. $\square$

Exercise 31 As in Example 3.2.6, let

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

Find a vector $\vec{x} \in \mathbb{R}^2$ so that $A\vec{x} \neq B\vec{x}$. [**Hint:** See Exercise 30.]

Despite some unexpected results such as in Example 3.2.5 and Example 3.2.6, the next theorem shows that the matrix–vector product behaves well with respect to matrix addition, vector addition and scalar multiplication, and follows some very familiar rules.

Theorem 3.2.8 **(Properties of the Matrix–Vector Product)**

Let $A, B \in M_{m\times n}(\mathbb{R})$, $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then

(a) $A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y}$.

(b) $A(c\vec{x}) = c(A\vec{x}) = (cA)\vec{x}$.

(c) $(A+B)\vec{x} = A\vec{x} + B\vec{x}$.

Proof: We prove (a). Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$ where $\vec{a}_1, \ldots, \vec{a}_n \in \mathbb{R}^m$, $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Then

$$\begin{aligned} A(\vec{x}+\vec{y}) &= \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \\ &= (x_1+y_1)\vec{a}_1 + \cdots + (x_n+y_n)\vec{a}_n \\ &= x_1\vec{a}_1 + y_1\vec{a}_1 + \cdots + x_n\vec{a}_n + y_n\vec{a}_n \\ &= (x_1\vec{a}_1 + \cdots + x_n\vec{a}_n) + (y_1\vec{a}_1 + \cdots + y_n\vec{a}_n) \\ &= A\vec{x} + A\vec{y} \end{aligned}$$ $\square$

Another important property involving multiplication of real numbers is that for any $x \in \mathbb{R}$ we have $1x = x$. As a result, we call 1 the *multiplicative identity*. It is natural to ask if there is a matrix A such that $A\vec{x} = \vec{x}$ for every $\vec{x} \in \mathbb{R}^n$.

Definition 3.2.9
Identity Matrix

The $n \times n$ **identity matrix**, denoted by I_n (or $I_{n\times n}$ or just I if the size is clear) is the square matrix of size $n \times n$ with $(I_n)_{ii} = 1$ for $i = 1, 2, \ldots, n$ and zeros elsewhere.

Example 3.2.10

For instance,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Theorem 3.2.11 For every $\vec{x} \in \mathbb{R}^n$, we have $I_n\vec{x} = \vec{x}$.

Proof: Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$. Then

$$I_n\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n = \vec{x}.$$ $\square$

Note that $I_n\vec{x} = \vec{x}$ for every $\vec{x} \in \mathbb{R}^n$ is exactly why we call I_n the identity matrix. It is also why we require I_n to be an square matrix. If I were an $m \times n$ matrix with $m \neq n$ and $\vec{x} \in \mathbb{R}^n$, then $I\vec{x} \in \mathbb{R}^m \neq \mathbb{R}^n$ so $I\vec{x}$ could never be equal to $\vec{x}$.

We end this section by showing that dot products can be used to compute matrix–vector products. Consider

$$A = \begin{bmatrix} 1 & -1 & 6 \\ 0 & 2 & 1 \\ 4 & -3 & 2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

so that

$$A\vec{x} = 1\begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} + 1\begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} + 2\begin{bmatrix} 6 \\ 1 \\ 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1(1)+1(-1)+2(6) \\ 1(0)+1(2)+2(1) \\ 1(4)+1(-3)+2(2) \end{bmatrix}}_{\text{these look like dot products}} = \begin{bmatrix} 12 \\ 4 \\ 5 \end{bmatrix}.$$

If we let $\vec{r}_1, \vec{r}_2, \vec{r}_3 \in \mathbb{R}^3$ be such that

$$\vec{r}_1^{\,T} = \begin{bmatrix} 1 & -1 & 6 \end{bmatrix}, \quad \vec{r}_2^{\,T} = \begin{bmatrix} 0 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \vec{r}_3^{\,T} = \begin{bmatrix} 4 & -3 & 2 \end{bmatrix}$$

are the rows of A, then we see from the above that the entries of $A\vec{x}$ are the dot products

$$\vec{r}_1 \cdot \vec{x} = \begin{bmatrix} 1 \\ -1 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \vec{r}_2 \cdot \vec{x} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{r}_3 \cdot \vec{x} = \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix},$$

that is

$$A\vec{x} = \begin{bmatrix} \vec{r}_1 \cdot \vec{x} \\ \vec{r}_2 \cdot \vec{x} \\ \vec{r}_3 \cdot \vec{x} \end{bmatrix}.$$

In general, given $A \in M_{m\times n}(\mathbb{R})$, there are vectors $\vec{r}_1, \ldots, \vec{r}_m \in \mathbb{R}^n$ so that

$$A = \begin{bmatrix} \vec{r}_1^{\,T} \\ \vdots \\ \vec{r}_m^{\,T} \end{bmatrix}$$

and for any $\vec{x} \in \mathbb{R}^n$,

$$A\vec{x} = \begin{bmatrix} \vec{r}_1^{\,T} \\ \vdots \\ \vec{r}_m^{\,T} \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{r}_1 \cdot \vec{x} \\ \vdots \\ \vec{r}_m \cdot \vec{x} \end{bmatrix}.$$

Thus, the ith entry of $A\vec{x}$ is the dot product $\vec{r}_i \cdot \vec{x}$ where $\vec{r}_i^{\,T}$ is the ith row of A.

Example 3.2.12 Let

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -4 \\ 3 & -1 \\ 7 & 2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Compute $A\vec{x}$.

Solution: We let

$$\vec{r}_1^{\,T} = \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad \vec{r}_2^{\,T} = \begin{bmatrix} 2 & -4 \end{bmatrix}, \quad \vec{r}_3^{\,T} = \begin{bmatrix} 3 & -1 \end{bmatrix} \quad \text{and} \quad \vec{r}_4^{\,T} = \begin{bmatrix} 7 & 2 \end{bmatrix}.$$

Then

$$A\vec{x} = \begin{bmatrix} \vec{r}_1^{\,T} \\ \vec{r}_2^{\,T} \\ \vec{r}_3^{\,T} \\ \vec{r}_4^{\,T} \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{r}_1 \cdot \vec{x} \\ \vec{r}_2 \cdot \vec{x} \\ \vec{r}_3 \cdot \vec{x} \\ \vec{r}_4 \cdot \vec{x} \end{bmatrix} = \begin{bmatrix} 1(1) - 1(2) \\ 1(2) - 1(-4) \\ 1(3) - 1(-1) \\ 1(7) - 1(2) \end{bmatrix} = \begin{bmatrix} -1 \\ 6 \\ 4 \\ 5 \end{bmatrix}.$$

The previous example seems like a lot of writing, but in practice we will only be computing the matrix–vector product for "small" matrices where we can perform the computations in our heads. Thus, it's okay to simply write

$$\begin{bmatrix} 1 & 2 \\ 2 & -4 \\ 3 & -1 \\ 7 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 6 \\ 4 \\ 5 \end{bmatrix}.$$

Exercise 32 Let

$$A = \begin{bmatrix} 1 & 1 & 2 & -1 \\ 2 & 1 & -3 & 2 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}.$$

Compute $A\vec{x}$ in two ways:

(a) Using the definition of the matrix–vector product.

(b) Using dot products.

Section 3.2 Problems

3.2.1. If possible, compute the following matrix-vector products in two ways: by using the definition of the matrix-vector product, and by using dot products. If not possible, explain why.

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

(b) $\begin{bmatrix} 1 & 2 & 2 & 1 \\ 2 & 3 & 3 & 4 \\ 1 & 2 & 3 & -3 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.

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

3.2.2. Given $A = \left[\begin{smallmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{smallmatrix}\right]$, $\vec{x} = \left[\begin{smallmatrix} 2 \\ 5 \\ 1 \end{smallmatrix}\right]$ and $\vec{b} = \left[\begin{smallmatrix} 3 \\ 9 \\ 12 \end{smallmatrix}\right]$. Verify that $A\vec{x} = \vec{b}$ and use this fact to write $\vec{b}$ as a linear combination of the columns of A.

3.2.3. Let A be the zero $m \times n$ matrix. Show that $A\vec{x} = \vec{0}$ for all $\vec{x} \in \mathbb{R}^n$.

3.2.4. (a) Disprove the following statement concerning $A \in M_{m\times n}(\mathbb{R})$ and $\vec{x} \in \mathbb{R}^n$.

If $A\vec{x} = \vec{0}$, then either A is the zero matrix or $\vec{x}$ is the zero vector.

(b) Prove the following statement concerning $A \in M_{m\times n}(\mathbb{R})$.

If $A\vec{x} = \vec{0}$ *for all* $\vec{x} \in \mathbb{R}^n$, then A is the zero matrix.

[**Hint:** See Exercise 30.]

3.2.5. Let $A, B \in M_{n\times n}(\mathbb{R})$ and $\vec{x} \in \mathbb{R}^n$. Prove that $(A + B)\vec{x} = A\vec{x} + B\vec{x}$.

3.2.6. Let $A \in M_{n\times n}(\mathbb{R})$ and suppose that $A\vec{x}_1 = \vec{0}$ and $A\vec{x}_2 = \vec{0}$. Prove that if $\vec{x}$ is a linear combination of $\vec{x}_1$ and $\vec{x}_2$ then $A\vec{x} = \vec{0}$.

3.3 The Matrix Equation $A\vec{x} = \vec{b}$

We now return to our study of systems of linear equations. The simplest linear equation is

$$ax = b$$

which has one solution if $a \neq 0$ (namely, $x = \frac{b}{a}$), infinitely many solutions if $a = b = 0$ (namely any $x \in \mathbb{R}$) and no solutions if $a = 0$ and $b \neq 0$. Our goal in this section is to show that the matrix–vector product can be used to express any system of linear equations in a similar way. By doing so, we will have a way to more deeply understand systems of linear equations.

Example 3.3.1

Let

$$A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & -4 & 3 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Compute $A\vec{x}$.

Solution: Using dot products, we have

$$A\vec{x} = \begin{bmatrix} 1 & 3 & -2 \\ -1 & -4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 - 2x_3 \\ -x_1 - 4x_3 + 3x_4 \end{bmatrix}.$$

Note that in Example 3.3.1, $A\vec{x} \in \mathbb{R}^2$ and that each entry of $A\vec{x}$ looks like the "left side" of a linear equation. Thus, if we consider a vector $\vec{b} \in \mathbb{R}^2$, say $\vec{b} = \left[\begin{smallmatrix} -7 \\ 8 \end{smallmatrix}\right]$, then equating $A\vec{x} = \vec{b}$ gives

$$\begin{bmatrix} x_1 + 3x_2 - 2x_3 \\ -x_1 - 4x_2 + 3x_3 \end{bmatrix} = \begin{bmatrix} -7 \\ 8 \end{bmatrix}$$

and equating entries gives the system of linear equations

$$\begin{array}{rcrcrcr} x_1 & + & 3x_2 & - & 2x_3 & = & -7 \\ -x_1 & - & 4x_2 & + & 3x_3 & = & 8 \end{array}.$$

We can see now that A is the coefficient matrix of this system while $\vec{b}$ is the constant vector. This idea extends naturally to any system of linear equations and thus motivates the following definition.

Definition 3.3.2
Matrix Equation, Vector of Variables

For a system of m linear equations in the n variables $x_1, \ldots, x_n$, with coefficient matrix $A \in M_{m\times n}(\mathbb{R})$ and constant vector $\vec{b} \in \mathbb{R}^m$, the equation

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

is called the **matrix equation of the system**. Here $\vec{x} = \left[\begin{smallmatrix} x_1 \\ \vdots \\ x_n \end{smallmatrix}\right] \in \mathbb{R}^n$ is the **vector of variables** of the system of linear equations.

Example 3.3.3 The matrix equation of the system of linear equations

$$\begin{array}{rcrcrcrcr} x_1 & - & x_2 & - & 2x_3 & + & x_4 & = & 1 \\ 2x_1 & - & 4x_2 & + & x_3 & - & 2x_4 & = & 2 \\ 5x_1 & + & 4x_2 & + & 4x_3 & + & 2x_4 & = & 5 \end{array}$$

is

$$\underbrace{\begin{bmatrix} 1 & -1 & -2 & 1 \\ 2 & -4 & 1 & -2 \\ 5 & 4 & 4 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}}_{\vec{b}}.$$

Example 3.3.4 Let

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 3 & -1 \\ -1 & -2 & 1 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Write out the system of linear equations that the matrix equation $A\vec{x} = \vec{b}$ represents.

Solution: We have that

$$A\vec{x} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 3 & -1 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 - x_3 \\ 2x_1 + 3x_2 - x_3 \\ -x_1 - 2x_2 + x_3 \end{bmatrix},$$

so $A\vec{x} = \vec{b}$ can be written as

$$\begin{bmatrix} x_1 + 3x_2 - x_3 \\ 2x_1 + 3x_2 - x_3 \\ -x_1 - 2x_2 + x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Thus the system of equations is

$$\begin{array}{rcrcrcr} x_1 & + & 3x_2 & - & x_3 & = & 1 \\ 3x_1 & + & 3x_2 & - & x_3 & = & 1 \\ -x_1 & - & 2x_2 & + & x_3 & = & 0 \end{array}.$$

Exercise 33 Write out the system of linear equations represented by the matrix equation $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 3 & -1 \\ 2 & 2 \\ -4 & 0 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 6 \\ 3 \\ 2 \\ 7 \end{bmatrix}.$$

The matrix equation $A\vec{x} = \vec{b}$ is more than just a compact way of representing a system of linear equations. Returning to Example 3.3.3, notice that the vector $\vec{s} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ is a solution to the system of equations given there. At the same time, we see that $\vec{x} = \vec{s}$ satisfies the corresponding matrix equation—that is, $A\vec{s} = \vec{b}$.

In general, if $A\vec{x} = \vec{b}$ is the matrix equation of a system of linear equations, then any vector $\vec{s}$ that satisfies this equation (meaning: $A\vec{s} = \vec{b}$) will satisfy the system of equations. Indeed, the entries of $A\vec{x}$ are the "left sides" of the system of equations and the entries of $\vec{b}$ are the "right sides." So if plugging in $\vec{x} = \vec{s}$ into $A\vec{x}$ equates it to $\vec{b}$, then it follows that the left sides and right sides of the system are equal, and hence that $\vec{s}$ is a solution to the system. This motivates the following definition.

Definition 3.3.5
Solution to Matrix Equation

Let $A \in M_{m\times n}(\mathbb{R})$ and $\vec{b} \in \mathbb{R}^m$. A vector $\vec{s} \in \mathbb{R}^n$ is a **solution** to the matrix equation $A\vec{x} = \vec{b}$ if $A\vec{s} = \vec{b}$.

From our discussion above, we see that $\vec{s}$ is a solution to the matrix equation $A\vec{x} = \vec{b}$ if and only if $\vec{x} = \vec{s}$ is a solution to the system of linear equations that underlies the matrix equation. The upshot is that we can now view systems of linear equations and their corresponding matrix form $A\vec{x} = \vec{b}$ as being one and the same. In particular, solving a system of equations amounts to "solving" the matrix equation $A\vec{x} = \vec{b}$—that is, finding vectors $\vec{s}$ such that $\vec{x} = \vec{s}$ satisfies the matrix equation.

Example 3.3.6

Let $A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 3 & 2 & 3 & 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 10 \\ 16 \end{bmatrix}$. Show that $\vec{x} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$ is a solution to $A\vec{x} = \vec{b}$.

Solution: Since

$$A\vec{x} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 3 & 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 16 \end{bmatrix} = \vec{b},$$

we conclude that $\vec{x} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$ is a solution to $A\vec{x} = \vec{b}$.

Note that Example 3.3.6 shows that $\vec{x} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$ is a solution to the system of equations

$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & + & x_3 & + & x_4 & = & 10 \\ 3x_1 & + & 2x_2 & + & 3x_3 & + & x_4 & = & 16 \end{array}.$$

Indeed, substituting $x_1 = 1$, $x_2 = 3$, $x_3 = 2$ and $x_4 = 1$ into this system will lead to the exact same computations required in Example 3.3.6.

Example 3.3.7

Let

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

In Example 3.2.5, we saw that

$$A\vec{x} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 2 \end{bmatrix} - 1\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Considering the equation $A\vec{x} = \vec{0}$, the above computation shows that $\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ is a solution to the homogeneous system of equations

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & 0 \\ 2x_1 & + & 2x_2 & = & 0 \end{array}.$$

Observe a matrix equation of the form $A\vec{x} = \vec{0}$, where the right-side is the zero vector, indicates that we are considering a homogeneous system of linear equations. To showcase the power of working with matrix equations, let us generalize a result that we obtained in Chapter 2: in Example 2.4.6, we proved that a linear combination of two solutions to a homogeneous system will again be a solution to that system. Below, we state a more general version of Example 2.4.6 and prove it using a matrix equations. Note how much simpler the algebra becomes!

Example 3.3.8

Consider the homogeneous system of equations $A\vec{x} = \vec{0}$ where $A \in M_{m\times n}(\mathbb{R})$ and $\vec{x} \in \mathbb{R}^n$. Assume $\vec{x}_1, \ldots, \vec{x}_k \in \mathbb{R}^n$ are solutions to this system and let $c_1, \ldots, c_k \in \mathbb{R}$. Show that $c_1\vec{x}_1 + \cdots + c_k\vec{x}_k$ is also a solution to $A\vec{x} = \vec{0}$.

Proof: Since $\vec{x}_1, \ldots, \vec{x}_k$ are solutions to $A\vec{x} = \vec{0}$, we have that $A\vec{x}_1 = \cdots = A\vec{x}_k = \vec{0}$. Then

$$\begin{aligned} A(c_1\vec{x}_1 + \cdots + c_k\vec{x}_k) &= A(c_1\vec{x}_1) + \cdots + A(c_k\vec{x}_k) && \text{by Theorem 3.2.8(a)} \\ &= c_1A\vec{x}_1 + \cdots + c_kA\vec{x}_k && \text{by Theorem 3.2.8(b)} \\ &= c_1\vec{0} + \cdots + c_k\vec{0} \\ &= \vec{0}. \end{aligned}$$

Thus $c_1\vec{x}_1 + \cdots + c_k\vec{x}_k$ is a solution to $A\vec{x} = \vec{0}$. □

Example 2.4.6 and Example 3.3.8 show that the set of solutions of a homogeneous solution are *closed under linear combinations*, that is, given k solutions to a homogeneous system of linear equations, any linear combination of those solutions will also be a solution to the homogeneous system. Sets that are closed under linear combinations will be explored more in Chapter 4.

Exercise 34

Let $A \in M_{m\times n}(\mathbb{R})$ and $\vec{b} \in \mathbb{R}^m$. Show that if $\vec{x}_1$ and $\vec{x}_2$ are solutions to $A\vec{x} = \vec{b}$, then $c\vec{x}_1 + (1-c)\vec{x}_2$ is also a solution to $A\vec{x} = \vec{b}$ for any $c \in \mathbb{R}$.

We close this section by using the matrix equation to gain new insight into systems of linear equations.

Consider again

$$A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & -4 & 3 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} -7 \\ 8 \end{bmatrix}.$$

and define

$$\vec{a}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} 3 \\ -4 \end{bmatrix} \quad \text{and} \quad \vec{a}_3 = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$$

so that $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix}$. We have seen that the matrix equation $A\vec{x} = \vec{b}$ represents the system of linear equations

$$\begin{array}{rcrcrcr} x_1 & + & 3x_2 & - & 2x_3 & = & -7 \\ -x_1 & - & 4x_2 & + & 3x_3 & = & 8. \end{array} \tag{3.2}$$

Now, if we evaluate $A\vec{x}$ using Definition 3.2.2, we obtain

$$\vec{b} = A\vec{x} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1\vec{a}_1 + x_2\vec{a}_2 + x_3\vec{a}_3.$$

From this, we see that $\vec{x}$ is a solution to (3.2) if and only if $\vec{b}$ can be expressed as a linear combination of the columns of A. Note that in this case, the coefficients that are used to express $\vec{b}$ as a linear combination of the columns of A are exactly the values of the variables that comprise the solution to (3.2). This leads to the following theorem, whose proof is similar to the derivation above and is thus omitted.

Theorem 3.3.9 Let $A \in M_{m\times n}(\mathbb{R})$ and $\vec{b} \in \mathbb{R}^m$. Then

(a) The system $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b}$ can be expressed as a linear combination of the columns of A.

(b) If $\vec{a}_1, \dots, \vec{a}_n$ are the columns of A and $\vec{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$, then $\vec{x} = \vec{s}$ satisfies $A\vec{x} = \vec{b}$ if and only if $s_1\vec{a}_1 + \dots + s_n\vec{a}_n = \vec{b}$.

Example 3.3.10 Let

$$A = \begin{bmatrix} 1 & 3 \\ -1 & -4 \\ 4 & 1 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 5 \\ -6 \\ 9 \end{bmatrix}.$$

(a) Show that $\vec{s} = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ is a solution to the matrix equation $A\vec{x} = \vec{b}$.

(b) Express $\vec{b}$ as a linear combination of the columns of A.

Solution:

(a) Since

$$A\vec{s} = \begin{bmatrix} 1 & 3 \\ -1 & -4 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ -6 \\ 9 \end{bmatrix} = \vec{b},$$

we see that $\vec{s} = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ is a solution to $A\vec{x} = \vec{b}$.

(b) From Theorem 3.3.9(b), we have that

$$\vec{b} = \begin{bmatrix} 5 \\ -6 \\ 9 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ -4 \\ 1 \end{bmatrix}.$$

Exercise 35 Let

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & 0 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

(a) Show that $\vec{s} = \left[\begin{smallmatrix} 1 \\ 0 \\ 2 \end{smallmatrix}\right]$ is a solution to the matrix equation $A\vec{x} = \vec{b}$.

(b) Express $\vec{b}$ as a linear combination of the columns of A.

Recall that when we first encountered linear combinations in Section 1.2, we noticed that when trying to write a vector as a linear combination of some given vectors, we wound up with a system of linear equations that we needed to solve. Theorem 3.3.9 confirms this, and also shows that every system of equations $A\vec{x} = \vec{b}$ can be viewed as checking if $\vec{b}$ can be expressed as a linear combination of the columns of A. This relationship will be useful in Chapter 4.

Section 3.3 Problems

3.3.1. Consider the system of equations

$$\begin{array}{rcrcrcr} x_1 & & & + & 2x_3 & = & 3 \\ x_1 & + & x_2 & + & x_3 & = & -2 \\ 4x_1 & - & 3x_2 & + & 12x_3 & = & 1 \end{array}$$

(a) Give the matrix A and the vectors $\vec{x}$ and $\vec{b}$ so that the above system can be expressed in the form $A\vec{x} = \vec{b}$.

(b) Solve the above of system of equations

(c) Using your work in parts (a) and (b) above, express $\vec{b}$ as a linear combination of the columns of A.

3.3.2. Let $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix} \in M_{m\times 3}(\mathbb{R})$ and $\vec{b} \in \mathbb{R}^m$. Show that if the system $A\vec{x} = \vec{b}$ has a solution, then $\vec{b} = s_1\vec{a}_1 + s_2\vec{a}_2 + s_3\vec{a}_3$ for some $s_1, s_2, s_3 \in \mathbb{R}$.

3.3.3. Let A be an $m \times n$ matrix, $\vec{x} \in \mathbb{R}^n$ and $\vec{b} \in \mathbb{R}^m$ with $\vec{b} \neq \vec{0}$. The equation $A\vec{x} = \vec{b}$ represents a non-homogeneous system of m equations in n variables. The system $A\vec{x} = \vec{0}$ is the corresponding homogeneous system. Let $\vec{y} \in \mathbb{R}^n$ satisfy the non-homogeneous system and $\vec{z} \in \mathbb{R}^n$ satisfy the corresponding homogeneous system.

(a) Show that the vector $\vec{y} + t\vec{z}$ satisfies the non-homogeneous system for any scalar t.

(b) Find all scalars s so that the vector $s\vec{y} + \vec{z}$ satisfies the non-homogeneous system.

(c) Find all scalars s so that the vector $s\vec{y} + \vec{z}$ satisfies the corresponding homogeneous system.

3.3.4. Let $A \in M_{m\times n}(\mathbb{R})$ and let $\vec{b} \in \mathbb{R}^m$ be a nonzero vector. Suppose that $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$ are solutions to the homogeneous system $A\vec{x} = \vec{0}$ and that $\vec{w}$ is a solution to the non-homogeneous system $A\vec{x} = \vec{b}$. Prove that $\vec{u} = \vec{w} + c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$ is a solution to the system $A\vec{x} = \vec{b}$ for any $c_1, \ldots, c_k \in \mathbb{R}$.

3.3.5. Let $A \in M_{m\times n}(\mathbb{R})$ be such that the system $A\vec{x} = \vec{e}_i$ is consistent for $i = 1, \ldots, m$. Show that $A\vec{x} = \vec{c}$ is consistent for every $\vec{c} \in \mathbb{R}^m$.

3.4 Matrix Multiplication

We now extend the matrix–vector product to matrix multiplication.

Definition 3.4.1
Matrix Product

If $A \in M_{m\times n}(\mathbb{R})$ and $B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_k \end{bmatrix} \in M_{n\times k}(\mathbb{R})$, then the **matrix product** AB is the $m \times k$ matrix

$$AB = \begin{bmatrix} A\vec{b}_1 & \cdots & A\vec{b}_k \end{bmatrix}.$$

That is, the columns of AB are the matrix–vector products $A\vec{b}_1, \ldots, A\vec{b}_k$.

Thus when computing the product AB for $A \in M_{m\times n}(\mathbb{R})$ and $B \in M_{n\times k}(\mathbb{R})$, we are computing n matrix–vector products. To understand why the product $AB \in M_{m\times k}(\mathbb{R})$ in Definition 3.4.1, note that since $B \in M_{n\times k}(\mathbb{R})$, each column $\vec{b}_i$ of B is a vector in $\mathbb{R}^n$. Thus the matrix–vector product $A\vec{b}_i \in \mathbb{R}^m$.

As with the matrix–vector product, the size of the matrices we are multiplying is important. It can help to remember the following:

$$\underset{\mathbf{m}\times\underbrace{\mathbf{n}\quad\mathbf{n}}_{\text{must agree}}\times\mathbf{k}}{A \qquad B} = \underset{\mathbf{m}\times\mathbf{k}}{AB}.$$

Example 3.4.2

Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}.$$

Then $A \in M_{2\times 2}(\mathbb{R})$ and $B \in M_{3\times 3}(\mathbb{R})$. Since the number of columns of A is not equal to the number of rows of B, the product AB is not defined, and since the number of columns of B is not equal to the number of rows of A, the product BA is not defined.

Thus, we have an example of matrices A and B where

$$AB \text{ and } BA \text{ are both undefined.}$$

Example 3.4.3

Let

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

Then the product AB is defined since $A \in M_{2\times 3}(\mathbb{R})$ has 3 columns and $B \in M_{3\times 2}(\mathbb{R})$ has 3 rows. The columns of B are

$$\vec{b}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{b}_2 = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}.$$

Since

$$A\vec{b}_1 = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 0 \end{bmatrix} \quad \text{and} \quad A\vec{b}_2 = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

we have that

$$AB = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 \end{bmatrix} = \begin{bmatrix} 9 & 6 \\ 0 & 1 \end{bmatrix}.$$

Exercise 36 Let

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

Compute AB.

The above method to multiply matrices can be quite tedious. As with the matrix–vector product, we can simplify the task using dot products. From Section 3.2, recall that for

$$A = \begin{bmatrix} \vec{r}_1^{\,T} \\ \vdots \\ \vec{r}_m^{\,T} \end{bmatrix} \in M_{m\times n}(\mathbb{R}) \quad \text{and} \quad \vec{x} \in \mathbb{R}^n$$

where $\vec{r}_1, \ldots, \vec{r}_m \in \mathbb{R}^n$, we have that

$$A\vec{x} = \begin{bmatrix} \vec{r}_1 \cdot \vec{x} \\ \vdots \\ \vec{r}_m \cdot \vec{x} \end{bmatrix}.$$

Thus for $B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_k \end{bmatrix} \in M_{n\times k}(\mathbb{R})$,

$$AB = A\begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_k \end{bmatrix} = \begin{bmatrix} A\vec{b}_1 & \cdots & A\vec{b}_k \end{bmatrix} = \begin{bmatrix} \vec{r}_1 \cdot \vec{b}_1 & \cdots & \vec{r}_1 \cdot \vec{b}_k \\ \vdots & & \vdots \\ \vec{r}_m \cdot \vec{b}_1 & \cdots & \vec{r}_m \cdot \vec{b}_k \end{bmatrix} \tag{3.3}$$

from which we see that the (i, j)-entry of AB is $\vec{r}_i \cdot \vec{b}_j$.

Example 3.4.4 As in Example 3.4.3, let

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

Then

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 2 & 3 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & -1 \\ 2 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1(1) + 2(1) + 3(2) & 1(2) + 2(-1) + 3(2) \\ -1(1) - 1(1) + 1(2) & -1(2) - 1(-1) + 1(2) \end{bmatrix} \\ &= \begin{bmatrix} 9 & 6 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

For matrices of small size, we normally evaluate products by performing the dot product calculations in our head. Thus in Example 3.4.4, it is okay to simply write

$$AB = \begin{bmatrix} 9 & 6 \\ 0 & 1 \end{bmatrix}$$

and not include the intermediate step.

Exercise 37 Let

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

Compute AB.

Next, we turn our attention to the algebraic properties of matrix multiplication. The following two examples demonstrate the important fact that *matrix multiplication is not commutative!*

Example 3.4.5 Let

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

Then

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 \\ 4 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 9 & -3 & 5 \\ 19 & -5 & 13 \end{bmatrix}.$$

Also note that $A \in M_{2\times2}(\mathbb{R})$ and $B \in M_{2\times3}(\mathbb{R})$ so $AB \in M_{2\times3}(\mathbb{R})$. However, the number of columns of B is not equal to the number of rows of A, so the product BA is not defined.

Thus, we have an example of matrices A and B where

$$AB \text{ is defined while } BA \text{ is undefined.}$$

Example 3.4.6 Let

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

Then

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

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

from which we see that

$$AB \neq BA$$

despite the products AB and BA both being defined and having the same size.

We learn from Examples 3.4.5 and 3.4.6 that, given two matrices A and B such that AB is defined, the product BA may not be defined, and even if it is, BA may not be equal to AB.

Exercise 38 Give an example of matrices A and B for which the products AB and BA are both defined but are of different sizes.

Exercise 39 Let

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad \text{and} \quad A_5 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}.$$

For each $i = 1, \ldots, 5$ and each $j = 1, \ldots, 5$, compute A_iA_j and A_jA_i whenever possible.

Recall that the transpose of a matrix was introduced in Section 3.1 and was used in Section 3.2 to give an efficient way to compute matrix–vector products using dot products. We give an example that shows that the transpose behaves oddly with matrix multiplication.

Example 3.4.7 Let

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

Then

$$(AB)^T = \left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \right)^T = \begin{bmatrix} -1 & 5 \\ -1 & 11 \end{bmatrix}^T = \begin{bmatrix} -1 & -1 \\ 5 & 11 \end{bmatrix}$$

but

$$A^TB^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 5 \\ 6 & 6 \end{bmatrix} \neq (AB)^T.$$

However

$$B^TA^T = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 5 & 11 \end{bmatrix} = (AB)^T.$$

Despite some peculiar behaviour, matrix multiplication does satisfy a lot of the familiar properties we know from multiplication of real numbers, as can be seen in the next theorem.

Theorem 3.4.8 **(Properties of Matrix Multiplication)**

Let $c \in \mathbb{R}$ and A, B, C be matrices of appropriate sizes. Then:

(a) $IA = A$. — I is an identity matrix

(b) $AI = A$. — I is an identity matrix

(c) $A(BC) = (AB)C$. — Matrix multiplication is associative

(d) $A(B + C) = AB + AC$. — Left distributive law

(e) $(B+C)A = BA + CA$. Right distributive law

(f) $(cA)B = c(AB) = A(cB)$.

(g) $(AB)^T = B^T A^T$.

Note that since we defined matrix products in terms of the matrix vector product, we have that (c) holds for the matrix vector product also: $A(B\vec{x}) = (AB)\vec{x}$ where $\vec{x}$ has the same number of entries as B has columns. We also note that (g) can be generalized as

$$(A_1 A_2 \cdots A_k)^T = A_k^T \cdots A_2^T A_1^T \tag{3.4}$$

where $A_1, \ldots, A_k$ are matrices of appropriate sizes.

Example 3.4.9 Simplify $A(3B-C) + (A-2B)C + 2B(C+2A)$.

Solution: We have

$$\begin{aligned} A(3B-C) + (A-2B)C + 2B(C+2A) &= 3AB - AC + AC - 2BC + 2BC + 4BA \\ &= 3AB + 4BA \end{aligned}$$

Make careful note of the following points regarding Example 3.4.9 – we must keep the order of our matrices correct when doing matrix algebra:

- $A(3B-C) = 3AB - AC$, that is, when distributing, A *must* remain on the left,
- $(A-2B)C = AC - 2BC$, that is, when distributing, C *must* remain on the right,
- $3AB + 4BA \neq 7AB$ since we *cannot* assume $AB = BA$.

Exercise 40 We say that two matrices $A_1, A_2 \in M_{n\times n}(\mathbb{R})$ *commute* if $A_1A_2 = A_2A_1$.

Assume that $A, B, C \in M_{n\times n}(\mathbb{R})$ are such that C commutes with both A and B. Show that C commutes with AB.

We end this section on matrix multiplication by defining powers of a square matrix.

Definition 3.4.10
Powers of a Matrix

Let $A \in M_{n\times n}(\mathbb{R})$. We define $A^2 = AA$ and for any integer $k \geq 2$, we define $A^k = AA^{k-1}$.

Note that powers of a non-square matrix are not defined since the product AA is not defined if the number of columns of A is not the same as the number of rows.

Example 3.4.11 Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$. Compute A^2, A^3 and A^4.

Solution: We have

$$
\begin{aligned}
A^2 &= \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix} \\
A^3 = AA^2 &= \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 9 & 10 \\ 10 & 4 \end{bmatrix} \\
A^4 = AA^3 &= \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 9 & 10 \\ 10 & 4 \end{bmatrix} = \begin{bmatrix} 29 & 18 \\ 18 & 20 \end{bmatrix}.
\end{aligned}
$$

Being able to compute powers of a matrix efficiently turns out to be an important aspect of many practical applications of linear algebra. As the above example demonstrates, computing A^k using the definition is quite tedious. For instance, to compute A^{10}, we need to compute A^9 first, which in turn needs A^8, A^7, and so on. We will later learn of a more efficient way of performing these computations in Chapter 8 The next exercise gives a preview.

Exercise 41 The matrix

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

is an example of a *diagonal matrix*, that is, a matrix whose non-zero entries occur only on the *main diagonal*. Compute D^{10}.

[**Hint:** If you compute D^2 and D^3, a pattern should emerge.]

Section 3.4 Problems

3.4.1. For each pair of matrices A and B, compute AB and BA, or explain why the products are not defined.

(a) $A = \begin{bmatrix} 1 & 3 & 3 \\ -2 & 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 2 \\ 4 & 1 \\ 1 & 1 \end{bmatrix}.$

(b) $A = \begin{bmatrix} -2 & 2 \\ 3 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}.$

(c) $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & -3 \\ 1 & -3 & 18 \end{bmatrix}, \quad B = \begin{bmatrix} 27 & -21 & -5 \\ -21 & 17 & 4 \\ -5 & 4 & 1 \end{bmatrix}.$

(d) $A = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, \quad B = \begin{bmatrix} -3 & 1 & 2 \end{bmatrix}.$

3.4.2. Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 2 & -1 \end{bmatrix}$. Find all real matrices B so that $AB = I_{2\times 2}$.

3.4.3. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$. Find all $B \in M_{2\times 2}(\mathbb{R})$ so that $BA = AB$.

3.4.4. All of the following statements are **FALSE**. In each case, give an example disproving the statement.

(a) For all $A, B \in M_{n\times n}(\mathbb{R})$, $(A+B)^2 = A^2 + 2AB + B^2$.

(b) If $A \in M_{n\times n}(\mathbb{R})$ and if $A^2 = 0_{n\times n}$, then $A = 0_{n\times n}$.

(c) If $A \in M_{n\times n}(\mathbb{R})$ and if $A^2 = I_n$, then either $A = I_n$ or $A = -I_n$.

3.4.5. Let $A, B \in M_{n\times n}(\mathbb{R})$.

(a) Show that $A(B\vec{x}) = (AB)\vec{x}$ for every $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$.

[**Hint**: Let $B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_n \end{bmatrix}$ and use the definitions of the matrix–vector product and matrix multiplication.]

(b) Show that $A(BC) = (AB)C$ for every $C \in M_{n\times n}(\mathbb{R})$.
[**Hint**: Let $C = \begin{bmatrix} \vec{c}_1 & \cdots & \vec{c}_n \end{bmatrix}$ and use the definition of matrix multiplication - you will need to use the result from part(a) at some point.]

3.4.6. Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{m\times n}(\mathbb{R})$. Prove that if A^TA is a zero matrix, then A is a zero matrix.

3.4.7. Let $A \in M_{m\times n}(\mathbb{R})$ and $B \in M_{n\times k}(\mathbb{R})$.

(a) Show that if B has a column of zeros, then so too does AB.

(b) Show that if A has a row of zeros, then so too does AB. [**Hint:** For a quick proof, take the transpose and use part (a).]

3.5 Matrix Inverses

We have seen that like real numbers, we can multiply appropriately sized matrices. For real numbers, we know that 1 is the multiplicative identity since $1(x) = x = x(1)$ for any $x \in \mathbb{R}$. We also know that if $x, y \in \mathbb{R}$ are such that $xy = 1 = yx$, then x and y are multiplicative inverses of each other, and we say that they are both invertible. We have recently seen that for an $n \times n$ matrix A, $IA = A = AI$ where I is the $n \times n$ identity matrix which shows that I is the multiplicative identity for $M_{n\times n}(\mathbb{R})$. It is then natural to ask that for a given a matrix A, does there exist a matrix B so that $AB = I = BA$? If so, the requirement that $AB = BA$ imposes the condition that A and B be a square matrices.

Definition 3.5.1
Invertible Matrix, Inverse Matrix

Let $A \in M_{n\times n}(\mathbb{R})$. If there exists a $B \in M_{n\times n}(\mathbb{R})$ such that

$$AB = I = BA$$

then A is **invertible** and B is an **inverse** of A (and B is invertible with A an inverse of B).

Note that our definition called B *an* inverse of A, instead of *the* inverse since it's not immediately clear whether or not there are multiple inverses for a given invertible matrix. In actuality, if $A \in M_{n\times n}(\mathbb{R})$ is invertible, then it has *exactly one* inverse. To see this, suppose that $B, C \in M_{n\times n}(\mathbb{R})$ are inverses of A. Then $BA = I$ and $AC = I$, and therefore

$$B = BI = B(AC) = (BA)C = IC = C.$$

So an invertible matrix has a uniquely determined inverse. In particular, we can speak of *the* inverse of A, and the following definition is unambiguous.

Definition 3.5.2
A^{-1}

If $A \in M_{n\times n}(\mathbb{R})$ is invertible, then we denote its inverse by A^{-1}.

Example 3.5.3

Let

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

Then

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

and

$$BA = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

so A is invertible and B is the inverse of A. Thus we can write

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

Example 3.5.4 Let

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

Then for any $b_1, b_2, b_3, b_4 \in \mathbb{R}$,

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} b_1 + 2b_3 & b_2 + 2b_4 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

so A is not invertible.

Notice that in the previous example, A is a nonzero matrix that fails to be invertible. This might be surprising since for a real number x, we know that x being invertible is equivalent to x being nonzero. Clearly this is not the case for $n \times n$ matrices.

By the above definition, to show that $B \in M_{n\times n}(\mathbb{R})$ is the inverse of $A \in M_{n\times n}(\mathbb{R})$, we must check that both $AB = I$ and $BA = I$. Then next theorem shows that if $AB = I$, then it follows that $BA = I$ (or equivalently, if $BA = I$ then it follows that $AB = I$) so that we need only verify only one of $AB = I$ and $BA = I$ to conclude that B is the inverse of A.

Theorem 3.5.5 Let $A, B \in M_{n\times n}(\mathbb{R})$ be such that $AB = I$. Then $BA = I$. Moreover, $\operatorname{rank}(A) = \operatorname{rank}(B) = n$.

Proof: Let $A, B \in M_{n\times n}(\mathbb{R})$ be such that $AB = I$. We first show that $\operatorname{rank}(B) = n$. Let $\vec{x} \in \mathbb{R}^n$ be such that $B\vec{x} = \vec{0}$. Since $AB = I$,

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

so $\vec{x} = \vec{0}$ is the only solution to the homogeneous system $B\vec{x} = \vec{0}$. Thus, $\operatorname{rank}(B) = n$ by part (b) of the System–Rank Theorem.

We next show that $BA = I$. Let $\vec{y} \in \mathbb{R}^n$. Since $\operatorname{rank}(B) = n$ and B has n rows, part (c) of the System–Rank Theorem guarantees that we will find $\vec{x} \in \mathbb{R}^n$ such that $\vec{y} = B\vec{x}$. Then

$$(BA)\vec{y} = (BA)B\vec{x} = B(AB)\vec{x} = BI\vec{x} = B\vec{x} = \vec{y} = I\vec{y}$$

so $(BA)\vec{y} = I\vec{y}$ for every $\vec{y} \in \mathbb{R}^n$. Thus $BA = I$ by the Matrix Equality Theorem .

Finally, since $BA = I$, it follows that $\operatorname{rank}(A) = n$ by the first part of our proof with the roles of A and B interchanged. □

We have now proven that if $A \in M_{n\times n}(\mathbb{R})$ is invertible, then $\operatorname{rank}(A) = n$. It follows that the reduced row echelon form of A is I.

Theorem 3.5.6 **(Properties of Matrix Inverses)**

Let $A, B \in M_{n\times n}(\mathbb{R})$ be invertible and let $c \in \mathbb{R}$ with $c \neq 0$. Then:

(a) $(cA)^{-1} = \frac{1}{c}A^{-1}$.

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

(c) $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$ for k a positive integer.

(d) $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

(e) $\left(A^{-1}\right)^{-1} = A$.

Proof: We prove (b) and (d) only. For (b), since

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

we have that $(AB)^{-1} = B^{-1}A^{-1}$ and for (d), since

$$A^T\left(A^{-1}\right)^T = \left(A^{-1}A\right)^T = I^T = I$$

we see that $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$. □

Exercise 42 Prove parts (a) and (e) of Theorem 3.5.6. [**Hint:** Mimic the proofs of parts (b) and (d) given above.]

Note that Theorem 3.5.6(b) generalizes for more than two matrices. For invertible matrices $A_1, A_2, \ldots, A_k \in M_{n\times n}(\mathbb{R})$ we have that $A_1A_2\cdots A_k$ is invertible and

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}\cdots A_2^{-1}A_1^{-1}.$$

In particular, if $A_1 = A_2 = \cdots = A_k = A$ is invertible, then

$$\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$$

for any positive integer k.

Example 3.5.7 Let A, B and C be invertible matrices of appropriate sizes. Express $\left(2AB^2C^T\right)^{-1}$ in terms of A^{-1}, B^{-1} and C^{-1}.

Solution: Using Theorem 3.5.6, we have

$$\begin{aligned}
\left(2AB^2C^T\right)^{-1} &= \left(C^T\right)^{-1}\left(B^2\right)^{-1}(2A)^{-1} && \text{by (b)}\\
&= \left(C^{-1}\right)^T\left(B^{-1}\right)^2\left(\frac{1}{2}A^{-1}\right) && \text{by (a), (c), (d)}\\
&= \frac{1}{2}\left(C^{-1}\right)^T\left(B^{-1}\right)^2A^{-1}.
\end{aligned}$$

3.5.1 Matrix Inversion Algorithm

Having shown many properties of matrix inverses, we have yet to actually compute the inverse of an invertible matrix. We know that a real number x is invertible if and only if $x \neq 0$, and in this case, $x^{-1} = \frac{1}{x}$. Things aren't quite so easy with matrices.[2] We derive an algorithm here that will tell us if a matrix is invertible, and compute the inverse should the matrix be invertible. Our construction is for 3×3 matrices, but generalizes naturally for $n \times n$ matrices.

Let $A \in M_{3\times 3}(\mathbb{R})$ and for a matrix $X = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \vec{x}_3 \end{bmatrix} \in M_{3\times 3}(\mathbb{R})$, consider the equation $AX = I$. Then

$$\begin{aligned} A\begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \vec{x}_3 \end{bmatrix} &= \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \end{bmatrix} \\ \begin{bmatrix} A\vec{x}_1 & A\vec{x}_2 & A\vec{x}_3 \end{bmatrix} &= \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \end{bmatrix}. \end{aligned}$$

Thus

$$A\vec{x}_1 = \vec{e}_1, \quad A\vec{x}_2 = \vec{e}_2 \quad \text{and} \quad A\vec{x}_3 = \vec{e}_3,$$

so we have three systems of equations, all with the same coefficient matrix. We consider two cases:

Case I: The RREF of A is I. In this case $\operatorname{rank}(A) = 3$, and since A is a 3×3 matrix, the system $A\vec{x}_1 = \vec{e}_1$ is consistent by the System–Rank Theorem (c) and has a unique solution $\vec{x}_1 = \vec{b}_1 \in \mathbb{R}^3$ by the System–Rank Theorem (b). Similarly, the systems $A\vec{x}_2 = \vec{e}_2$ and $A\vec{x}_3 = \vec{e}_3$ are consistent with unique solutions $\vec{x}_2 = \vec{b}_2$ and $\vec{x}_3 = \vec{b}_3 \in \mathbb{R}^3$. We define $B = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{bmatrix} \in M_{3\times 3}(\mathbb{R})$. Then

$$AX = AB = A\begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{bmatrix} = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 & A\vec{b}_3 \end{bmatrix} = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \end{bmatrix} = I,$$

so A is invertible and $A^{-1} = B$.

Case II: The RREF of A is not I. Then $\operatorname{rank}(A) < 3$ and A cannot be invertible since if A were invertible, we would have $\operatorname{rank}(A) = 3$ by Theorem 3.5.5.

Our above derivation will require us to solve the three systems of linear equations

$$A\vec{x}_1 = \vec{e}_1, \quad A\vec{x}_2 = \vec{e}_2 \quad \text{and} \quad A\vec{x}_3 = \vec{e}_3$$

so we will have to consider three augmented matrices

$$\left[A \mid \vec{e}_1\right], \quad \left[A \mid \vec{e}_2\right] \quad \text{and} \quad \left[A \mid \vec{e}_3\right].$$

Row reducing the first of these augmented matrices will inform us as to whether or not A is invertible. Assuming A is invertible, then we will find that

$$\left[A \mid \vec{e}_1\right] \longrightarrow \left[I \mid \vec{b}_1\right].$$

for some unique $\vec{b}_1 \in \mathbb{R}^3$. We will then need to solve the other two systems as well to find that

$$\left[A \mid \vec{e}_2\right] \longrightarrow \left[I \mid \vec{b}_2\right] \quad \text{and} \quad \left[A \mid \vec{e}_3\right] \longrightarrow \left[I \mid \vec{b}_3\right]$$

[2]Don't even think about writing $A^{-1} = \frac{1}{A}$. This makes no sense as $\frac{1}{A}$ is not even defined.

for some unique $\vec{b}_2, \vec{b}_3 \in \mathbb{R}^3$. Notice that the exact same elementary row operations will be performed to reduce all three of these augmented matrices. Thus, we solve all three systems at once by considering the *super–augmented* matrix

$$\left[A \mid \vec{e}_1 \quad \vec{e}_2 \quad \vec{e}_3\right] = \left[A \mid I\right].$$

Again, under these same row operations,

$$\left[A \mid I\right] \longrightarrow \left[I \mid \vec{b}_1 \quad \vec{b}_2 \quad \vec{b}_3\right] = \left[I \mid B\right]$$

and we will have $A^{-1} = B$.

The same method works for $n \times n$ matrices. We summarize our observations below.

ALGORITHM (Matrix Inversion)

To determine if $A \in M_{n\times n}(\mathbb{R})$ is invertible (and to find A^{-1} if it exists), perform the following steps.

- **Step 1:** Row reduce $\left[A \mid I_n\right]$ to RREF: $\left[A \mid I_n\right] \to \left[R \mid B\right]$. (Note that R is the reduced row echelon form of A.)
- **Step 2:** Refer to R and B in **Step 1**.
  - If $R = I_n$, then A is invertible, and $A^{-1} = B$.
  - If $R \neq I_n$, then A is not invertible.

Example 3.5.8 Let

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}.$$

Find A^{-1} if it exists.

Solution: We have

$$\left[\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 4 & 5 & 0 & 1 \end{array}\right] \underset{R_2-2R_1}{\longrightarrow} \left[\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 0 & -1 & -2 & 1 \end{array}\right] \overset{R_1+3R_2}{\longrightarrow} \left[\begin{array}{cc|cc} 2 & 0 & -5 & 3 \\ 0 & -1 & -2 & 1 \end{array}\right] \begin{array}{c} \frac{1}{2}R_1 \\ -R_2 \end{array} \left[\begin{array}{cc|cc} 1 & 0 & -5/2 & 3/2 \\ 0 & 1 & 2 & -1 \end{array}\right].$$

So A is invertible (since the reduced row echelon form of A is I) and

$$A^{-1} = \begin{bmatrix} -5/2 & 3/2 \\ 2 & -1 \end{bmatrix}.$$

Example 3.5.9 Let

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

Find A^{-1} if it exists.

Solution: We have

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array}\right] \underset{R_2-2R_1}{\longrightarrow} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array}\right]$$

We see that the reduced row echelon form of A is

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so A is not invertible (note that $\operatorname{rank}(A) = 1 < 2$).

Exercise 43 Let

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

Find A^{-1} if it exists.

Note that if you find A to be invertible and you compute A^{-1}, then you can check your work by ensuring that $AA^{-1} = I$.

3.5.2 Properties of Matrix Inverses

Theorem 3.5.10 **(Cancellation Laws)**

Let $A \in M_{n\times n}(\mathbb{R})$ be invertible.

(a) For all $B, C \in M_{n\times k}(\mathbb{R})$, if $AB = AC$, then $B = C$. left cancellation

(b) For all $B, C \in M_{k\times n}(\mathbb{R})$, if $BA = CA$, then $B = C$. right cancellation

Proof: We prove (a). We have

$$\begin{aligned} AB &= AC \\ A^{-1}(AB) &= A^{-1}(AC) \\ (A^{-1}A)B &= (A^{-1}A)C \\ IB &= IC \\ B &= C. \end{aligned}$$

□

Note that our two cancellation laws require that A be invertible. Indeed

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \end{bmatrix}$$

but
$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \neq \begin{bmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \end{bmatrix}.$$
Notice that $\operatorname{rank}\left(\left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]\right) = 1 < 2$ so $\left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ is not invertible.

Example 3.5.11 If $A, B, C \in M_{n\times n}(\mathbb{R})$ are such that A is invertible and $AB = CA$, does $B = C$?

Solution: The answer is no. To see this, consider
$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}.$$
Then
$$AB = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$$
and
$$CA = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix},$$
so $AB = CA$ but $B \neq C$.

The previous example shows that we do not have *mixed cancellation*. This is a direct result of matrix multiplication not being commutative. From $AB = CA$ with A invertible, we can obtain $B = A^{-1}CA$, and since $B \neq C$, we have $C \neq A^{-1}CA$. Note that we cannot cannot cancel A and A^{-1} here.

Example 3.5.12 For $A, B \in M_{n\times n}(\mathbb{R})$ with A, B and $A + B$ invertible, do we have that $(A + B)^{-1} = A^{-1} + B^{-1}$?

Solution: The answer is no. Let $A = B = I$. Then $A + B = 2I$ and
$$(A + B)^{-1} = (2I)^{-1} = \frac{1}{2}I^{-1} = \frac{1}{2}I$$
but
$$A^{-1} + B^{-1} = I^{-1} + I^{-1} = I + I = 2I$$
As $\frac{1}{2}I \neq 2I$, $(A + B)^{-1} \neq A^{-1} + B^{-1}$.

Exercise 44 Give an example of two non-invertible matrices A and B such that $A + B$ is invertible.

The following theorem summarizes many of the results we have seen thus far in the course, and shows the importance of matrix invertibility. This theorem is central to all of linear algebra and actually contains many more parts, some of which we will encounter later. Note that we have already proven all of these equivalences.

Theorem 3.5.13 **(Matrix Invertibility Criteria)**

Let $A \in M_{n\times n}(\mathbb{R})$. The following are equivalent. (That is, the following statements are either *all true* or *all false*.)

(a) A is invertible.

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

(c) The reduced row echelon form of A is I.

(d) For all $\vec{b} \in \mathbb{R}^n$, the system $A\vec{x} = \vec{b}$ is consistent and has a unique solution.

(e) A^T is invertible.

Exercise 45 Prove that if A is invertible, then properties (b), (c), (d) and (e) of Theorem 3.5.13 are true. [**Hint:** If you are stuck, review the earlier parts of the notes. The proofs all occur somewhere!]

In particular, for A invertible, the system $A\vec{x} = \vec{b}$ has a unique solution. We can solve for $\vec{x}$ using our matrix algebra:

$$\begin{aligned} A\vec{x} &= \vec{b} \\ A^{-1}A\vec{x} &= A^{-1}\vec{b} \\ I\vec{x} &= A^{-1}\vec{b} \\ \vec{x} &= A^{-1}\vec{b}. \end{aligned}$$

Example 3.5.14 Consider the system of equations $A\vec{x} = \vec{b}$ with

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}.$$

Then A is invertible (see Example 3.5.8) and

$$\begin{aligned} \vec{x} &= A^{-1}\vec{b} \\ &= \begin{bmatrix} -5/2 & 3/2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} -23/2 \\ 9 \end{bmatrix} \end{aligned}$$

Of course we could have solved the above system $A\vec{x} = \vec{b}$ by row reducing the augmented matrix $\left[A \mid \vec{b}\right] \to \left[I \mid {}^{-23/2}_{\ \ 9}\right]$. Note that to find A^{-1}, we row reduced $\left[A \mid I\right] \longrightarrow \left[I \mid A^{-1}\right]$, and that the elementary row operations used in both cases are the same.

Section 3.5 Problems

3.5.1. Use the matrix inversion algorithm to find the inverses of the following matrices, if possible.

(a) $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. (b) $B = \begin{bmatrix} -2 & -1 \\ -1 & -1 \end{bmatrix}$. (c) $C = \begin{bmatrix} 2 & 8 \\ 1 & 4 \end{bmatrix}$.

(d) $D = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 5 & 3 \\ 0 & -3 & -2 \end{bmatrix}$. (e) $E = \begin{bmatrix} 3 & 2 & 6 \\ 2 & 3 & 5 \\ 1 & 1 & 2 \end{bmatrix}$. (f) $F = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$.

3.5.2. Find all values of $a \in \mathbb{R}$, if any, for which the matrix $A = \left[\begin{smallmatrix} a & 1 & 1 \\ 1 & a & 1 \\ 1 & 1 & a \end{smallmatrix}\right]$ is invertible.

3.5.3. Let $A = \left[\begin{smallmatrix} 2 & 5 & 1 \\ 1 & 2 & 0 \\ 4 & 5 & -2 \end{smallmatrix}\right]$, $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]$ and $\vec{b} = \left[\begin{smallmatrix} 10 \\ 4 \\ 9 \end{smallmatrix}\right]$ and consider the equation $A\vec{x} = \vec{b}$.

(a) Write out the system of equations represented by the equation $A\vec{x} = \vec{b}$.

(b) Use the matrix inversion algorithm to find A^{-1}. Verify that your answer is correct by showing that $A^{-1}A = I$.

(c) Use A^{-1} to find the solution to the system $A\vec{x} = \vec{b}$

(d) Using your answer from part (c), express $\vec{b}$ as a linear combination of the columns of A.

3.5.4. Find the 2×2 matrix A if

$$\left(\begin{bmatrix} 2 & -1 \\ 3 & -4 \end{bmatrix} + (2A)^T \right)^{-1} = \begin{bmatrix} 2 & 9 \\ 1 & 4 \end{bmatrix}.$$

3.5.5. Let $A \in M_{n\times n}(\mathbb{R})$ be an invertible matrix. Prove the following.

(a) $\left(A^{-1}\right)^{-1} = A$.

(b) $(cA)^{-1} = \frac{1}{c}A^{-1}$ for $c \in \mathbb{R}$, $c \neq 0$.

3.5.6. Let $A, B, C \in M_{n\times n}(\mathbb{R})$. Prove that if A and ABC are invertible, then B is invertible.

Chapter 4

Subspaces of $\mathbb{R}^n$

4.1 Spanning Sets

Recall that linear combinations were introduced in Section 1.2 where we observed that determining whether a vector could be expressed as a linear combination of some given vectors amounted to examining a linear system of equations. More recently, we encountered linear combinations in Section 3.2 were we learned that every linear combination could be expressed as a matrix–vector product. The present section explores linear combinations in more depth and will employ the matrix–vector equation and the System–Rank Theorem to help us generate some very useful results. Some background material involving sets will also be required, so it will be helpful to read through Appendix A to ensure you understand the basic set theoretic notions.

We begin by considering a set $S = \{\vec{v}_1, \dots, \vec{v}_k\}$ of k vectors in $\mathbb{R}^n$. The set of all linear combinations of these vectors will be of great interest to us. This motivates the following definition.

Definition 4.1.1
Span, Spanning Set

Let $S = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$. The **span** of S is the set

$$\operatorname{Span} S = \{c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \mid c_1, \dots, c_k \in \mathbb{R}\}.$$

We say that

- $\operatorname{Span} S$ is **spanned** by S, and
- S is a **spanning set** for $\operatorname{Span} S$, or that S **spans** $\operatorname{Span} S$.

By convention, we define $\operatorname{Span} \varnothing = \{\vec{0}\}$.

It follows from Definition 4.1.1 that in order to show that a vector $\vec{x} \in \mathbb{R}^n$ belongs to $\operatorname{Span} S$, we must show that we can express $\vec{x}$ as a linear combination of $\vec{v}_1, \dots, \vec{v}_k$. If we cannot do this, then $\vec{x} \notin \operatorname{Span} S$.

Example 4.1.2 Let $S = \left\{ \left[\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right] \right\}$. Then $\left[\begin{smallmatrix}2\\3\\0\end{smallmatrix}\right] \in \operatorname{Span} S$ because

$$\begin{bmatrix}2\\3\\0\end{bmatrix} = 2\begin{bmatrix}1\\0\\0\end{bmatrix} + 3\begin{bmatrix}0\\1\\0\end{bmatrix}.$$

On the other hand,

$$\begin{bmatrix}0\\0\\1\end{bmatrix} \notin \operatorname{Span} S$$

because every linear combination of the vectors in S will have a 0 in the third entry.

Exercise 46 Show that $\left[\begin{smallmatrix}a\\b\\0\end{smallmatrix}\right] \in \operatorname{Span}\left\{ \left[\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right] \right\}$ for all $a, b \in \mathbb{R}$.

It is important to note that Definition 4.1.1 mentions two distinct sets:

(1) The set $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$, which is simply a set containing the k vectors $\vec{v}_1, \ldots, \vec{v}_k$ in $\mathbb{R}^n$ (or S is the empty set).

(2) The set $\operatorname{Span} S$, which is the set of *all* linear combinations of the k vectors in S (or $\operatorname{Span} S$ is the set containing just the zero vector in the event that S is the empty set).

Example 4.1.3 For any set $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$, show that $S \subseteq \operatorname{Span} S$.

Solution: Let $\vec{x} \in S$. Then $\vec{x} = \vec{v}_i$ for some $i = 1, \ldots, k$. Since

$$\vec{v}_i = 0\vec{v}_1 + \cdots + 0\vec{v}_{i-1} + 1\vec{v}_i + 0\vec{v}_{i+1} + \cdots + 0\vec{v}_k$$

we see that $\vec{x} = \vec{v}_i \in \operatorname{Span} S$. Thus $S \subseteq \operatorname{Span} S$.

Exercise 47 Show that $\vec{0} \in \operatorname{Span} S$ for any set of vectors $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ in $\mathbb{R}^n$.

Since $\operatorname{Span} \varnothing = \{\vec{0}\}$, Example 4.1.3 and Exercise 47 hold for $S = \varnothing$ as well.

The previous examples can be solved by inspection once Definition 4.1.1 is understood. The following examples are more involved.

Example 4.1.4 Determine if

$$\begin{bmatrix}2\\3\end{bmatrix} \in \operatorname{Span}\left\{ \begin{bmatrix}4\\5\end{bmatrix}, \begin{bmatrix}3\\3\end{bmatrix} \right\}.$$

If so, express $\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]$ as a linear combination of $\left[\begin{smallmatrix}4\\5\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}3\\3\end{smallmatrix}\right]$.

Solution: To determine if $\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right] \in \operatorname{Span}\{\left[\begin{smallmatrix}4\\5\end{smallmatrix}\right], \left[\begin{smallmatrix}3\\3\end{smallmatrix}\right]\}$, we must determine if there are real numbers $c_1, c_2 \in \mathbb{R}$ such that

$$\begin{bmatrix}2\\3\end{bmatrix} = c_1\begin{bmatrix}4\\5\end{bmatrix} + c_2\begin{bmatrix}3\\3\end{bmatrix} = \begin{bmatrix}4c_1+3c_2\\5c_1+3c_2\end{bmatrix}.$$

By equating entries, we obtain the system of linear equations

$$\begin{array}{rcrcr} 4c_1 & + & 3c_2 & = & 2\\ 5c_1 & + & 3c_2 & = & 3 \end{array}.$$

Carrying the augmented matrix of this system to reduced row echelon form gives

$$\left[\begin{array}{cc|c}4&3&2\\5&3&3\end{array}\right] \xrightarrow{R_1\leftrightarrow R_2} \left[\begin{array}{cc|c}5&3&3\\4&3&2\end{array}\right] \xrightarrow{R_1-R_2} \left[\begin{array}{cc|c}1&0&1\\4&3&2\end{array}\right] \underset{R_2-4R_1}{\longrightarrow} \left[\begin{array}{cc|c}1&0&1\\0&3&-2\end{array}\right] \underset{\frac{1}{3}R_2}{\longrightarrow}$$
$$\left[\begin{array}{cc|c}1&0&1\\0&1&-2/3\end{array}\right].$$

As the system is consistent, we conclude that

$$\begin{bmatrix}2\\3\end{bmatrix} \in \operatorname{Span}\left\{\begin{bmatrix}4\\5\end{bmatrix}, \begin{bmatrix}3\\3\end{bmatrix}\right\}.$$

From the above reduced row echelon form, we see that $c_1 = 1$ and $c_2 = -\frac{2}{3}$. Thus

$$\begin{bmatrix}2\\3\end{bmatrix} = 1\begin{bmatrix}4\\5\end{bmatrix} - \frac{2}{3}\begin{bmatrix}3\\3\end{bmatrix}.$$

Example 4.1.5

Determine if

$$\begin{bmatrix}1\\2\\3\end{bmatrix} \in \operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}\right\}.$$

If so, express $\left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right]$ as a linear combination of $\left[\begin{smallmatrix}1\\0\\1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right]$.

Solution: Let $c_1, c_2 \in \mathbb{R}$ and consider

$$\begin{bmatrix}1\\2\\3\end{bmatrix} = c_1\begin{bmatrix}1\\0\\1\end{bmatrix} + c_2\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}c_1+c_2\\c_2\\c_1\end{bmatrix}.$$

We obtain the system of equations

$$\begin{array}{rcrcr} c_1 & + & c_2 & = & 1\\ & & c_2 & = & 2\\ c_1 & & & = & 3 \end{array}.$$

Solving this system, we have

$$\left[\begin{array}{cc|c}1&1&1\\0&1&2\\1&0&3\end{array}\right] \underset{R_3-R_1}{\longrightarrow} \left[\begin{array}{cc|c}1&1&1\\0&1&2\\0&-1&2\end{array}\right] \underset{R_3+R_2}{\longrightarrow} \left[\begin{array}{cc|c}1&1&1\\0&1&2\\0&0&4\end{array}\right],$$

which shows the system is inconsistent.

Here we see that $\begin{bmatrix}1\\2\\3\end{bmatrix}$ cannot be expressed as a linear combination of $\begin{bmatrix}1\\0\\1\end{bmatrix}$ and $\begin{bmatrix}1\\1\\0\end{bmatrix}$ and so we conclude that

$$\begin{bmatrix}1\\2\\3\end{bmatrix} \notin \operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}\right\}.$$

Example 4.1.6 Determine if

$$\begin{bmatrix}4\\7\\3\end{bmatrix} \in \operatorname{Span}\left\{\begin{bmatrix}1\\3\\1\end{bmatrix}, \begin{bmatrix}2\\1\\1\end{bmatrix}, \begin{bmatrix}3\\4\\2\end{bmatrix}\right\}$$

If so, express $\begin{bmatrix}4\\7\\3\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\3\\1\end{bmatrix}$, $\begin{bmatrix}2\\1\\1\end{bmatrix}$ and $\begin{bmatrix}3\\4\\2\end{bmatrix}$.

Solution: Let $c_1, c_2, c_3 \in \mathbb{R}$ and consider

$$\begin{bmatrix}4\\7\\3\end{bmatrix} = c_1\begin{bmatrix}1\\3\\1\end{bmatrix} + c_2\begin{bmatrix}2\\1\\1\end{bmatrix} + c_3\begin{bmatrix}3\\4\\2\end{bmatrix} = \begin{bmatrix}c_1+2c_2+3c_3\\3c_1+c_2+4c_3\\c_1+c_2+2c_3\end{bmatrix}.$$

We obtain the system of equations

$$\begin{array}{rcrcrcl} c_1 & + & 2c_2 & + & 3c_3 & = & 4\\ 3c_1 & + & c_2 & + & 4c_3 & = & 7\\ c_1 & + & c_2 & + & 2c_3 & = & 3 \end{array}.$$

Carrying the augmented matrix of this system to reduced row echelon form gives

$$\left[\begin{array}{ccc|c}1&2&3&4\\3&1&4&7\\1&1&2&3\end{array}\right] \underset{\substack{R_2-3R_1\\R_3-R_1}}{\longrightarrow} \left[\begin{array}{ccc|c}1&2&3&4\\0&-5&-5&-5\\0&-1&-1&-1\end{array}\right] \underset{-\frac{1}{5}R_2}{\longrightarrow} \left[\begin{array}{ccc|c}1&2&3&4\\0&1&1&1\\0&-1&-1&-1\end{array}\right] \overset{R_1-2R_2}{\underset{R_3+R_2}{\longrightarrow}} \left[\begin{array}{ccc|c}1&0&1&2\\0&1&1&1\\0&0&0&0\end{array}\right].$$

As the system is consistent, we conclude that

$$\begin{bmatrix}4\\7\\3\end{bmatrix} \in \operatorname{Span}\left\{\begin{bmatrix}1\\3\\1\end{bmatrix}, \begin{bmatrix}2\\1\\1\end{bmatrix}, \begin{bmatrix}3\\4\\2\end{bmatrix}\right\}$$

The solution to the system is $c_1 = 2-t$, $c_2 = 1-t$ and $c_3 = t$ where $t \in \mathbb{R}$. Taking $t = 0$ gives

$$\begin{bmatrix}4\\7\\3\end{bmatrix} = 2\begin{bmatrix}1\\3\\1\end{bmatrix} + 1\begin{bmatrix}2\\1\\1\end{bmatrix} + 0\begin{bmatrix}3\\4\\2\end{bmatrix}.$$

The existence of a parameter in our solution means that there are infinitely many ways to express $\begin{bmatrix}4\\7\\3\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\3\\1\end{bmatrix}$, $\begin{bmatrix}2\\1\\1\end{bmatrix}$ and $\begin{bmatrix}3\\4\\2\end{bmatrix}$.

For instance, we can take $t = 100$ to obtain

$$\begin{bmatrix} 4 \\ 7 \\ 3 \end{bmatrix} = -98 \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - 99 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + 100 \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}.$$

Exercise 48 Determine if

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

If so, express $\left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right]$ as a linear combination of $\left[\begin{smallmatrix} 1 \\ -1 \\ 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 3 \\ 0 \\ 1 \end{smallmatrix}\right]$.

As mentioned right after Definition 4.1.1 and observed in Example 4.1.4, Example 4.1.5 and Example 4.1.6, verifying if $\vec{v} \in \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\}$ amounts to determining if $\vec{v}$ can be expressed as a linear combination of $\vec{v}_1, \dots, \vec{v}_k$, that is, determining if there are $c_1, \dots, c_k \in \mathbb{R}$ so that

$$\vec{v} = c_1 \vec{v}_1 + \cdots + c_k \vec{v}_k.$$

Now recalling the matrix–vector product from Section 3.2, if we set $\vec{c} = \left[\begin{smallmatrix} c_1 \\ \vdots \\ c_k \end{smallmatrix}\right]$ and let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix}$ be the matrix whose columns are the vectors $\vec{v}_1, \dots, \vec{v}_k$, then we see that the above equation can be re-written as

$$\vec{v} = A\vec{c}.$$

Since we want to find a $\vec{c}$ that makes this equation hold, it follows that we need to determine if the system

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

is consistent. This verifies the following theorem, which is simply a restatement of Theorem 3.3.9(a) using Definition 4.1.1.

Theorem 4.1.7 Let $S = \{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^n$, $\vec{v} \in \mathbb{R}^n$ and let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix} \in M_{n \times k}(\mathbb{R})$. Then $\vec{v} \in \operatorname{Span} S$ if and only if the system $A\vec{x} = \vec{v}$ is consistent.

By Theorem 4.1.7, to check if $\vec{v} \in \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\}$, we need only verify that the system $A\vec{x} = \vec{v}$ is consistent, which amounts to carrying the augmented matrix of the system to row echelon form and applying part (a) of the System–Rank Theorem. However, if we wish to explicitly write $\vec{v}$ as a linear combination of $\vec{v}_1, \dots, \vec{v}_k$, then we must solve the system. In either case, we can simply start with the augmented matrix

$$\left[A \mid \vec{v} \right] = \left[\vec{v}_1 \ \cdots \ \vec{v}_k \mid \vec{v} \right].$$

Example 4.1.8 To illustrate this theorem, let's return to Examples 4.1.4 and 4.1.5.

To determine if $\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right] \in \operatorname{Span}\left\{\left[\begin{smallmatrix}4\\5\end{smallmatrix}\right], \left[\begin{smallmatrix}3\\3\end{smallmatrix}\right]\right\}$, we must check if the system

$$\begin{bmatrix}4 & 3\\5 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix}$$

is consistent by row reducing the augmented matrix

$$\left[\begin{array}{cc|c}4 & 3 & 2\\5 & 3 & 3\end{array}\right].$$

If you look at our work in Example 4.1.4, you will see that this is precisely what we did.

Likewise, to determine if $\left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right] \in \operatorname{Span}\left\{\left[\begin{smallmatrix}1\\0\\1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right]\right\}$, we must check if the system

$$\begin{bmatrix}1 & 1\\0 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix}$$

is consistent by row reducing the augmented matrix

$$\left[\begin{array}{cc|c}1 & 1 & 1\\0 & 1 & 2\\1 & 0 & 3\end{array}\right]$$

which is what we did in Example 4.1.5.

Exercise 49 Let $S = \{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^n$. Use Theorem 4.1.7 to show that $\vec{0} \in \operatorname{Span} S$.

(Compare with what you did in Exercise 47. Make sure to understand that these two solutions are essentially the same.)

Thus far, given a set $S = \{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^n$, we have been concerned with determining if a given vector $\vec{v} \in \mathbb{R}^n$ belongs to $\operatorname{Span} S$. It is natural to ask if *every* vector $\vec{v} \in \mathbb{R}^n$ belongs to $\operatorname{Span} S$, that is, if S spans $\mathbb{R}^n$.

In terms of sets, to show that S spans $\mathbb{R}^n$, we must show that $\operatorname{Span} S = \mathbb{R}^n$. According to Definition A.1.10, we must show that

(1) $\operatorname{Span} S \subseteq \mathbb{R}^n$, and

(2) $\mathbb{R}^n \subseteq \operatorname{Span} S$.

Note that since $S \subseteq \mathbb{R}^n$ and since $\mathbb{R}^n$ is closed under linear combinations by properties V1 and V4 of the Fundamental Properties of Vector Algebra, we have immediately that $\operatorname{Span} S \subseteq \mathbb{R}^n$. Thus we normally don't verify or even mention (1). Thus we simply need to verify (2) to show that $\operatorname{Span} S = \mathbb{R}^n$. It follows from Definition A.1.8 that we must pick an arbitrary $\vec{v} \in \mathbb{R}^n$ and show that $\vec{v} \in \operatorname{Span} S$. Theorem 4.1.10 shows that we can accomplish this by showing that $A\vec{x} = \vec{v}$ is consistent for *every* $\vec{v} \in \mathbb{R}^n$.

Let's look at an example.

Example 4.1.9 Let

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}.$$

Determine if $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ spans $\mathbb{R}^3$.

Solution: Let $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix}$. It follows from Theorem 4.1.7 that we must determine if the system $A\vec{x} = \vec{v}$ is consistent for every $\vec{v} \in \mathbb{R}^3$. However, part (c) of the System–Rank Theorem gives that $A\vec{x} = \vec{v}$ is consistent for every $\vec{v} \in \mathbb{R}^3$ if and only if $\operatorname{rank}(A) = 3$ (the number of rows of A). Thus we need only look at any row echelon form of A. We have

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 3 & 4 \end{bmatrix} \underset{\substack{R_2 - R_1 \\ R_3 - 2R_1}}{\longrightarrow} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \underset{R_3 - R_2}{\longrightarrow} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence $\operatorname{rank}(A) = 3$, so the system $A\vec{x} = \vec{v}$ is consistent for every $\vec{v} \in \mathbb{R}^3$, that is, S spans $\mathbb{R}^n$.

Note that the method used in Example 4.1.9 does not tell us *how* to express a vector $\vec{v} \in \mathbb{R}^3$ as a linear combination of the vectors in $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, just that it *can* be done for any $\vec{v} \in \mathbb{R}^3$. If we additionally need to know how to write $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \\ v_3 \end{smallmatrix}\right]$ as a linear combination of the vectors in S, we can carry the augmented matrix for $A\vec{x} = \vec{v}$ to reduced row echelon form:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & v_1 \\ 1 & 2 & 2 & v_2 \\ 2 & 3 & 4 & v_3 \end{array}\right] \underset{\substack{R_2 - R_1 \\ R_3 - 2R_1}}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 1 & 1 & v_1 \\ 0 & 1 & 1 & -v_1 + v_2 \\ 0 & 1 & 2 & -2v_1 + v_3 \end{array}\right] \underset{R_3 - R_2}{\overset{R_1 - R_2}{\longrightarrow}}$$
$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 2v_1 - v_2 \\ 0 & 1 & 1 & -v_1 + v_2 \\ 0 & 0 & 1 & -v_1 - v_2 + v_3 \end{array}\right] \underset{R_2 - R_3}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2v_1 - v_2 \\ 0 & 1 & 0 & 2v_2 - v_3 \\ 0 & 0 & 1 & -v_1 - v_2 + v_3 \end{array}\right].$$

We then have

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = (2v_1 - v_2) \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + (2v_2 - v_3) \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + (-v_1 - v_2 + v_3) \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}.$$

The next Theorem generalizes Example 4.1.9 to any set $S = \{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^n$.

Theorem 4.1.10 Let $S = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix} \in M_{n \times k}(\mathbb{R})$. Then S spans $\mathbb{R}^n$ if and only if $\operatorname{rank}(A) = n$.

Proof: It follows from Theorem 4.1.7 that that S spans $\mathbb{R}^n$ if and only if the system $A\vec{x} = \vec{v}$ is consistent for every $\vec{v} \in \mathbb{R}^n$, which is equivalent to $\operatorname{rank}(A) = n$ by part (c) of the System–Rank Theorem. $\square$

Example 4.1.11 Consider the set

$$S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

Since

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 3 & 2 \\ 1 & 1 & 1 \end{bmatrix} \underset{\substack{R_2 - 2R_1 \\ R_3 - R_1}}{\longrightarrow} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & -1 \end{bmatrix},$$

we see that $\operatorname{rank}(A) = 3$, so S spans $\mathbb{R}^3$ by Theorem 4.1.10.

Exercise 50 Determine if the set

$$S = \left\{ \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ 6 \end{bmatrix} \right\}$$

spans $\mathbb{R}^3$.

Example 4.1.12 Consider the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

from Example 4.1.5. Since

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \in M_{3\times 2}(\mathbb{R}),$$

we have that $\operatorname{rank}(A) \leq \min\{3, 2\} = 2 < 3$, so S does not span $\mathbb{R}^3$ by Theorem 4.1.10.

Note that in Example 4.1.12, we did not explicitly compute the rank of A, but instead used the fact that A had fewer columns than rows to show that the $\operatorname{rank}(A) < 3$. This, of course, is because $S \subseteq \mathbb{R}^3$ had fewer than 3 vectors. The following corollary[1] generalizes this observation.

Corollary 4.1.13 Let $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$. If $k < n$, then S cannot span $\mathbb{R}^n$.

Exercise 51 Prove Corollary 4.1.13.

It follows from Corollary 4.1.13 that if $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$ spans $\mathbb{R}^n$, then $k \geq n$, that is, we need at least n vectors to span $\mathbb{R}^n$. However, S having $k \geq n$ vectors does not guarantee that S spans $\mathbb{R}^n$.

Exercise 52 Give an example of a set $S \subseteq \mathbb{R}^3$ containing 4 vectors that does not span $\mathbb{R}^3$.

[1] A *corollary* is a result that immediately follows from a given theorem – in this case, Theorem 4.1.10.

Section 4.1 Problems

4.1.1. For each of the following, determine if $\vec{v} \in \operatorname{Span} S$. If so, express $\vec{v}$ as a linear combination of the vectors in S.

(a) $\vec{v} = \begin{bmatrix} 11 \\ 4 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 7 \\ 2 \end{bmatrix} \right\}$.

(b) $\vec{v} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 15 \\ 21 \end{bmatrix} \right\}$.

(c) $\vec{v} = \begin{bmatrix} 1 \\ -1 \\ 8 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 10 \end{bmatrix} \right\}$.

(d) $\vec{v} = \begin{bmatrix} -18 \\ -34 \\ 11 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ -2 \end{bmatrix}, \begin{bmatrix} 10 \\ 19 \\ -6 \end{bmatrix} \right\}$.

(e) $\vec{v} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$, $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Hint: To save some work, refer to Problem 2.2.1.

4.1.2. For each of the following, determine a condition that $v_1, v_2, v_3 \in \mathbb{R}$ must satisfy in order for $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \\ v_3 \end{smallmatrix}\right] \in \operatorname{Span} S$.

(a) $S = \left\{ \begin{bmatrix} 2 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 12 \\ 13 \\ -14 \end{bmatrix}, \begin{bmatrix} -8 \\ -6 \\ 4 \end{bmatrix} \right\}$.

(b) $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ 0 \\ 12 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \\ -2 \end{bmatrix} \right\}$.

4.1.3. (a) Determine if $S = \left\{ \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ is a spanning set for $\mathbb{R}^2$.

(b) Determine if $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \right\}$ is a spanning set for $\mathbb{R}^3$.

(c) Determine if $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a spanning set for $\mathbb{R}^3$.

(d) Determine if $S = \left\{ \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 7 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \right\}$ is a spanning set for $\mathbb{R}^3$.

4.1.4. Let $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \subseteq \mathbb{R}^3$ and let $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{bmatrix} \in M_{3\times 3}(\mathbb{R})$. Show that if A is invertible, then S spans $\mathbb{R}^3$.

4.1.5. Let $S = \{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^n$.

(a) Prove that if $\vec{w} \in \mathbb{R}^n$ is orthogonal to each of $\vec{v}_1, \dots, \vec{v}_k$, then $\vec{w}$ is orthogonal to every $\vec{x} \in \operatorname{Span} S$.

(b) For $A \in M_{m\times n}(\mathbb{R})$, prove that if $A\vec{v}_i = \vec{0}$ for $i = 1, \dots, n$, then $A\vec{x} = \vec{0}$ for every $\vec{x} \in \operatorname{Span} S$.

4.2 Geometry of Spanning Sets

Consider a set $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$ and define $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix} \in M_{n \times k}(\mathbb{R})$. For $\vec{v} \in \mathbb{R}^n$, Theorem 4.1.7 showed us that $\vec{v} \in \operatorname{Span} S$ if and only if the system $A\vec{x} = \vec{v}$ is consistent. Theorem 4.1.10 then showed us that $\operatorname{Span} S = \mathbb{R}^n$ if and only if $\operatorname{rank}(A) = n$. However, we have little information about $\operatorname{Span} S$ if $\operatorname{rank}(A) < n$, aside from $\operatorname{Span} S \neq \mathbb{R}^n$. This section will further develop our geometric understanding of $\operatorname{Span} S$. We will focus on $\mathbb{R}^3$ and will first consider the set $\operatorname{Span}\{\vec{v}_1\}$, where $\vec{v}_1 \in \mathbb{R}^3$.

Example 4.2.1 Describe the subset $U = \operatorname{Span}\{\vec{0}\}$ of $\mathbb{R}^3$ geometrically.

Solution: Since

$$U = \operatorname{Span}\{\vec{0}\} = \{c_1\vec{0} \mid c_1 \in \mathbb{R}\} = \{\vec{0}\},$$

U is the origin of $\mathbb{R}^3$.

Example 4.2.2 Describe the subset

$$U = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

of $\mathbb{R}^3$ geometrically.

Solution: By definition,

$$U = \left\{ c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \,\middle|\, c_1 \in \mathbb{R} \right\}.$$

Thus, $\vec{x} \in U$ if and only if it satisfies

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \tag{4.1}$$

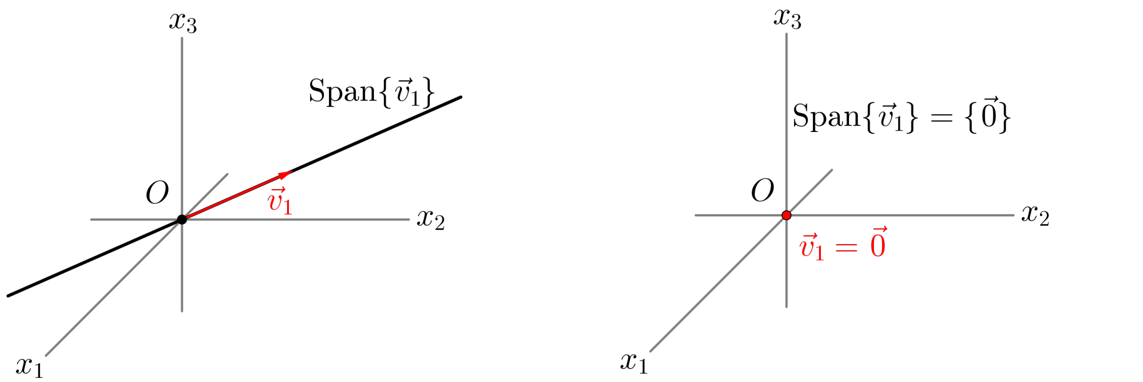
for some $c_1 \in \mathbb{R}$. We recognize (4.1) as a vector equation for a line. Hence, U is a line in $\mathbb{R}^3$ through the origin.

Exercise 53 Let $\vec{v}_1 \in \mathbb{R}^3$ be any nonzero vector. Show that $U = \operatorname{Span}\{\vec{v}_1\}$ is a line through the origin.

From Example 4.2.1 and Exercise 53, we see that for $\vec{v}_1 \in \mathbb{R}^3$, $\operatorname{Span}\{\vec{v}_1\}$ is either

- the origin of $\mathbb{R}^3$ if $\vec{v}_1 = \vec{0}$, or
- a line through the origin in $\mathbb{R}^3$ if $\vec{v}_1 \neq \vec{0}$.

It follows that $U = \operatorname{Span}\{\vec{v}_1\}$ fails to be a line exactly when $\vec{v}_1 = \vec{0}$. This is illustrated in Figure 4.2.1.



(a) If $\vec{v}_1 \neq \vec{0}$, then Span$\{\vec{v}_1\}$ is a line through the origin with direction vector $\vec{v}_1$.

(b) If $\vec{v}_1 = \vec{0}$, then Span$\{\vec{v}_1\}$ is the set $\{\vec{0}\}$.

Figure 4.2.1: Geometrically interpreting Span$\{\vec{v}_1\}$ in $\mathbb{R}^3$. The picture in $\mathbb{R}^n$ is similar.

We now consider Span$\{\vec{v}_1, \vec{v}_2\}$ for $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^3$.

Example 4.2.3 Describe the subset

$$U_1 = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

of $\mathbb{R}^3$ geometrically.

Solution: By definition,

$$U_1 = \left\{ c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \;\middle|\; c_1, c_2 \in \mathbb{R} \right\}$$

so $\vec{x} \in U_1$ if and only if it satisfies

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \tag{4.2}$$

for some $c_1, c_2 \in \mathbb{R}$. Since neither $\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$ nor $\left[\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right]$ is a scalar multiple of the other, we recognize (4.2) as the vector equation of a plane. Hence U_1 is a plane in $\mathbb{R}^3$ through the origin.

As a side note, the set U in Example 4.2.3 is from Example 4.1.5. In light of what we have observed here, Example 4.1.5 shows us that the point $P(1, 2, 3)$ does not lie on the plane U.

We saw previously that for $\vec{v}_1 \in \mathbb{R}^3$, we are not guaranteed that Span$\{\vec{v}_1\}$ will be a line through the origin. We will now see that for $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^3$, Span$\{\vec{v}_1, \vec{v}_2\}$ will not always be plane through the origin in $\mathbb{R}^3$.

Example 4.2.4 Consider the subset

$$U_2 = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix} \right\}$$

of $\mathbb{R}^3$. By definition,

$$U_2 = \left\{ c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix} \,\middle|\, c_1, c_2 \in \mathbb{R} \right\}$$

so $\vec{x} \in U_2$ if and only if it satisfies

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix} \tag{4.3}$$

for some $c_1, c_2 \in \mathbb{R}$. We notice, however, that $\left[\begin{smallmatrix} -2 \\ 0 \\ -2 \end{smallmatrix}\right] = -2 \left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$, so $\vec{x} \in U_2$ if and only if

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \left(-2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) = (c_1 - 2c_2) \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Since every $\vec{x} \in U$ is a scalar multiple of the single vector $\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$, we see that U is not a plane in $\mathbb{R}^3$.

It is now entirely natural to ask what U_2 is geometrically. We know that every $\vec{x} \in U_2$ is of the form

$$\vec{x} = (c_1 - 2c_2) \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

for some $c_1, c_2 \in \mathbb{R}$. This looks suspiciously like the vector equation

$$\vec{x} = t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

which we know to be the vector equation of a line through the origin with direction vector $\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$. We can define this line as the set

$$L = \left\{ t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\} = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We now see that our work in Example 4.2.4 shows that $U_2 \subseteq L$. However, before we can say that U_2 is a line through the origin, we must show that $L \subseteq U_2$. To achieve this, we must show that every $\vec{y} \in L$ can be expressed as a linear combination of $\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} -2 \\ 0 \\ -2 \end{smallmatrix}\right]$.

Example 4.2.5 Let

$$U_2 = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix} \right\} \quad \text{and} \quad L = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Show that $U_2 = L$.

Solution: We must show that $U_2 \subseteq L$ and $L \subseteq U_2$. To show that $U_2 \subseteq L$, we pick $\vec{x} \in U_2$ and then show that $\vec{x} \in L$. Thus we let $\vec{x} \in U_2$. Then there are $c_1, c_2 \in \mathbb{R}$ so that

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix}.$$

Since $\left[\begin{smallmatrix} -2 \\ 0 \\ -2 \end{smallmatrix}\right] = -2 \left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$, we have that

$$\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix} = \vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \left(-2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right) = (c_1 - 2c_2) \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \in L,$$

which shows that $U_2 \subseteq L$. We now show that $L \subseteq U_2$. To do so, we let $\vec{y} \in L$ and show that $\vec{y} \in U_2$. Since $\vec{y} \in L$, there is $d_1 \in \mathbb{R}$ so that

$$\vec{y} = d_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = d_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} -2 \\ 0 \\ -2 \end{bmatrix} \in U_2,$$

so $L \subseteq U_2$. Since $U_2 \subseteq L$ and $L \subseteq U_2$, we have that $U_2 = L$, as required.

We have now shown that $U_2 = L$ and may conclude that the set U_2 presented in Example 4.2.4 is indeed a line through the origin. It's worth noting that the our work to verify that $U_2 \subseteq L$ in Example 4.2.5 is identical to our work in Example 4.2.4. As we continue to develop our geometric intuition about spanning sets, we will verify our observations by proving set equality as needed.

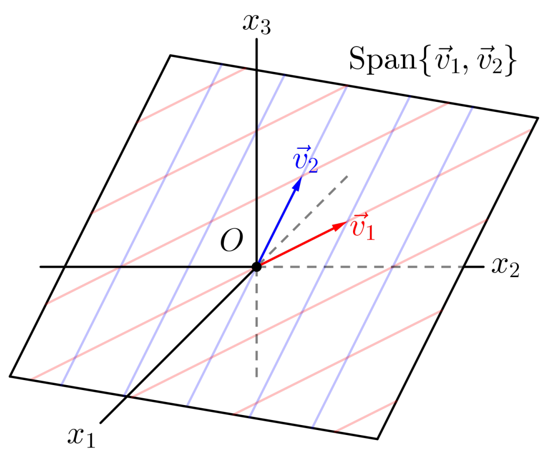
Recall that the spanning sets for U_1 and U_2 from Example 4.2.3 and Example 4.2.4 each contained two vectors, but that we obtained a plane in Example 4.2.3 and a line in Example 4.2.4. This is because in Example 4.2.4, one of the vectors in the spanning set for U_2 was a scalar multiple of the other and as a result, we could express one of the vectors in terms of the other. This *dependency* among the vectors in the spanning set for U_2 means that we can remove one of the vectors and the resulting set containing just one vector will still span U_2.

Exercise 54 Let $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^3$ and let $U = \text{Span}\{\vec{v}_1, \vec{v}_2\}$. Show that if $\vec{v}_2 = c\vec{v}_1$, then U is the origin if $\vec{v}_1 = \vec{0}$ and a line through the origin otherwise.

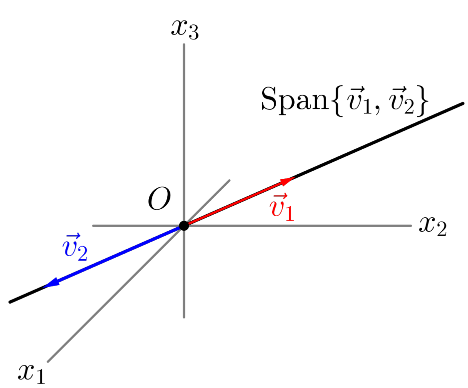
From Example 4.2.3, Example 4.2.4 and Exercise 54, we see that for $\vec{v}_1, \vec{v}_2 \in \mathbb{R}$, $U = \text{Span}\{\vec{v}_1, \vec{v}_2\}$ is

- a plane through the origin in $\mathbb{R}^3$ if neither $\vec{v}_1$ nor $\vec{v}_2$ is a scalar multiple of the other (loosely speaking, this means that $\vec{v}_1$ and $\vec{v}_2$ are both nonzero and don't "point" in the same or opposite directions), or
- a line through the origin in $\mathbb{R}^3$ if at least one of $\vec{v}_1$ and $\vec{v}_2$ is nonzero and a scalar multiple of the other (loosely speaking, either $\vec{v}_1$ and $\vec{v}_2$ are both nonzero and point in the same direction or opposite directions, or exactly one of $\vec{v}_1$ and $\vec{v}_2$ is the zero vector), or
- the origin of $\mathbb{R}^3$ if $\vec{v}_1 = \vec{v}_2 = \vec{0}$.

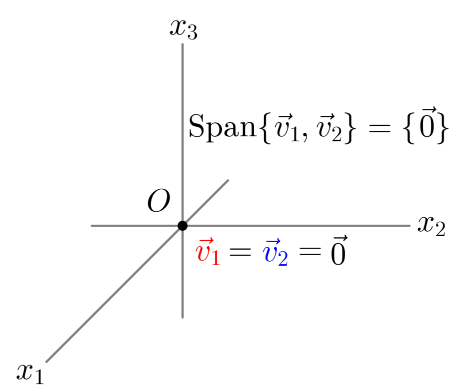
It follows that $U = \text{Span}\{\vec{v}_1, \vec{v}_2\}$ fails to be a plane exactly when at least one of $\vec{v}_1$ and $\vec{v}_2$ is a scalar multiple of the other. This is illustrated in Figure 4.2.2.



(a) If neither $\vec{v}_1$ nor $\vec{v}_2$ is a scalar multiples of the other, then Span$\{\vec{v}_1, \vec{v}_2\}$ is a plane through the origin.



(b) If at least one of $\vec{v}_1$ and $\vec{v}_2$ is nonzero and a scalar multiple of the other, then Span$\{\vec{v}_1, \vec{v}_2\}$ is a line through the origin. The direction vector of this line will be whichever of $\vec{v}_1$ and $\vec{v}_2$ is nonzero.



(c) If $\vec{v}_1 = \vec{v}_2 = \vec{0}$, then Span$\{\vec{v}_1, \vec{v}_2\}$ is simply the set $\{\vec{0}\}$.

Figure 4.2.2: Geometrically interpreting Span$\{\vec{v}_1, \vec{v}_2\}$ in $\mathbb{R}^3$. The picture in $\mathbb{R}^n$ is similar.

Note that it is more complicated to describe the span of a set of two vectors than it is to describe the span of a set of one vector. We now turn out attention to considering Span$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ for $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^3$.

Example 4.2.6 Let

$$U = \operatorname{Span}\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\2\\1\end{bmatrix} \right\}.$$

Show that $U = \mathbb{R}^3$.

Solution: Let

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

Then $\operatorname{rank}(A) = 3$. It follows from Theorem 4.1.10 that $U = \mathbb{R}^3$.

Exercise 55 Let

$$S = \left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\2\\1\end{bmatrix} \right\}.$$

Express $\vec{v} = \left[\begin{smallmatrix}v_1\\v_2\\v_3\end{smallmatrix}\right] \in \mathbb{R}^3$ as a linear combination of the vectors in S.

As with our examples with one and two vectors, things aren't always so simple.

Example 4.2.7 Let

$$U = \operatorname{Span}\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix} \right\} \quad \text{and} \quad V = \operatorname{Span}\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix} \right\}.$$

Show that $U = V$.

Solution: We will prove that $U \subseteq V$ and that $V \subseteq U$. We first show that $U \subseteq V$. Let $\vec{x} \in U$. Then for some $c_1, c_2, c_3 \in \mathbb{R}$,

$$\vec{x} = c_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + c_2 \begin{bmatrix}0\\1\\0\end{bmatrix} + c_3 \begin{bmatrix}1\\1\\0\end{bmatrix}.$$

However, we observe that $\left[\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right] = \left[\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right] + \left[\begin{smallmatrix}0\\1\\0\end{smallmatrix}\right]$ so

$$\begin{aligned}
\vec{x} &= c_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + c_2 \begin{bmatrix}0\\1\\0\end{bmatrix} + c_3 \left(\begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\1\\0\end{bmatrix} \right)\\
&= (c_1 + c_3) \begin{bmatrix}1\\0\\0\end{bmatrix} + (c_2 + c_3) \begin{bmatrix}0\\1\\0\end{bmatrix}\\
&\in \operatorname{Span}\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix} \right\} = V.
\end{aligned}$$

Thus $U \subseteq V$. Now let $\vec{y} \in V$. Then for some $d_1, d_2 \in \mathbb{R}$,

$$\begin{aligned}\vec{y} &= d_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + d_2 \begin{bmatrix}0\\1\\0\end{bmatrix}\\ &= d_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + d_2 \begin{bmatrix}0\\1\\0\end{bmatrix} + 0 \begin{bmatrix}1\\1\\0\end{bmatrix}\\ &\in \operatorname{Span}\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix} \right\} = U.\end{aligned}$$

Thus $V \subseteq U$. Hence $U = V$.

Let

$$S = \left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix} \right\} \quad \text{and} \quad C = \left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix} \right\}.$$

In Example 4.2.7, we were given $U = \operatorname{Span} S$, and we then showed that $U = \operatorname{Span} C$. Note that this is very similar to what we observed in Example 4.2.5: there was a *dependency* among the vectors in S (the given spanning set for U), that allowed us to express one of the vectors in S in terms of the remaining vectors in S. We saw we could then remove this vector from S to obtain the smaller set C which still spanned U . There is an important difference with Example 4.2.7, however: none of the vectors in S are a scalar multiple of any of the other vectors in S.

Thus, when trying to geometrically understand $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$, it is not enough to simply check whether there is a vector in $\{\vec{v}_1, \ldots, \vec{v}_k\}$ that is a scalar multiple of any of the other vectors, but rather, we must check whether any vector in $\{\vec{v}_1, \ldots, \vec{v}_k\}$ can be expressed as a *linear combination* of the other vectors. This is exhibited in Figure 4.2.3

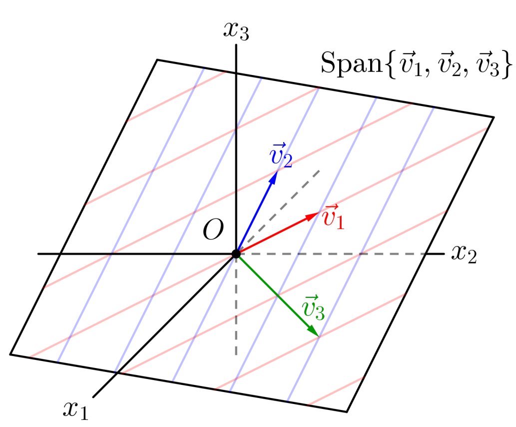


Figure 4.2.3: The set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ only spans a plane despite containing three vectors. Notice that none of $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$ are scalar multiples of any of the others. However, we see that $\vec{v}_3 = 2\vec{v}_1 - 2\vec{v}_2$, that is, that $\vec{v}_3$ can be expressed as a linear combination of $\vec{v}_1$ and $\vec{v}_2$.

As our goal for this section is to geometrically understand the span of a set of vectors, we have focused our attention in $\mathbb{R}^3$. We have noticed that given $\{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^3$, $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$ can be one of the following:

- $\{\vec{0}\}$,
- a line through the origin,
- a plane through the origin,
- all of $\mathbb{R}^3$.

Of course, we will observe similar outcomes in $\mathbb{R}^n$. For example, for $\{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^2$, we will find that $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$ will either be $\{\vec{0}\}$, a line through the origin, or all of $\mathbb{R}^2$.

We also observed in Example 4.2.7 that for $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^3$, if say, $\vec{v}_3$ could be expressed as a linear combination of $\vec{v}_1$ and $\vec{v}_2$, then $\operatorname{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$. The following theorem generalizes this for k vectors in $\mathbb{R}^n$.

Theorem 4.2.8 **(Reduction Theorem)**

Let $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$. One of these vectors, say $\vec{v}_i$, can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k$ if and only if

$$\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}.$$

We make a comment here before giving the proof. The statement we need to prove is a double implication, so we must prove the two implications:

(1) If $\vec{v}_i$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k$, then $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$

(2) If $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$, then $\vec{v}_i$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k$.

The result of this theorem is that the two statements

“ $\vec{v}_i$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k$”

and

“$\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$”

are equivalent, that is, they are both true or they are both false. The proof that follows is often not completely understood after just the first reading - it takes a bit of time to understand, so don't be discouraged if you need to read it a few times before it fully makes sense.

Proof of the Reduction Theorem: Without loss of generality[2], we assume $i = k$. To simplify the writing of the proof, we let

$$\begin{aligned} U &= \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}, \vec{v}_k\} \\ V &= \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}. \end{aligned}$$

To prove the first implication, assume that $\vec{v}_k$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_{k-1}$. Then there exist $c_1, \ldots, c_{k-1} \in \mathbb{R}$ such that

$$\vec{v}_k = c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1}. \tag{4.4}$$

We must show that $U = V$. Let $\vec{x} \in U$. Then there exist $d_1, \ldots, d_{k-1}, d_k \in \mathbb{R}$ such that

$$\vec{x} = d_1\vec{v}_1 + \cdots + d_{k-1}\vec{v}_{k-1} + d_k\vec{v}_k$$

and we make the substitution for $\vec{v}_k$ using (4.4) to obtain

$$\begin{aligned} \vec{x} &= d_1\vec{v}_1 + \cdots + d_{k-1}\vec{v}_{k-1} + d_k\big(c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1}\big) \\ &= (d_1 + d_kc_1)\vec{v}_1 + \cdots + (d_{k-1} + d_kc_{k-1})\vec{v}_{k-1} \end{aligned}$$

from which we see that $\vec{x}$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_{k-1}$ and it follows that $\vec{x} \in V$. Hence $U \subseteq V$. Now let $\vec{y} \in V$. Then there exist $a_1, \ldots, a_{k-1} \in \mathbb{R}$ such that

$$\begin{aligned} \vec{y} &= a_1\vec{v}_1 + \cdots + a_{k-1}\vec{v}_{k-1} \\ &= a_1\vec{v}_1 + \cdots + a_{k-1}\vec{v}_{k-1} + 0\vec{v}_k \end{aligned}$$

and we have that $\vec{y}$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_k$ from which it follows that $\vec{y} \in U$. We have that $V \subseteq U$ and combined with $U \subseteq V$, we conclude that $U = V$.

To prove the second implication, we now assume that $U = V$ and must show that $\vec{v}_k$ can be expressed as a linear combination of $\vec{v}_1, \ldots, \vec{v}_{k-1}$. Since $\vec{v}_k \in U$ (recall that $\vec{v}_k = 0\vec{v}_1 + \cdots + 0\vec{v}_{k-1} + 1\vec{v}_k$) and $U = V$, we have $\vec{v}_k \in V$. Thus, there exist $b_1, \ldots, b_{k-1} \in \mathbb{R}$ such that $\vec{v}_k = b_1\vec{v}_1 + \cdots + b_{k-1}\vec{v}_{k-1}$ as required. $\square$

The Reduction Theorem allows us to simplify spanning sets by removing "redundant" vectors (specifically, vectors that are linear combinations of other vectors in the spanning set). The next example illustrates this.

Example 4.2.9 Consider

$$U = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

Since $\left[\begin{smallmatrix} 5 \\ 0 \end{smallmatrix}\right] = 5\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] = 5\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] + 0\left[\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right] + 0\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$, the vector $\left[\begin{smallmatrix} 5 \\ 0 \end{smallmatrix}\right]$ is "redundant" and so the Reduction Theorem gives

$$U = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

---

[2]What we mean here is that if $i \neq k$, then we may "rename" the vectors $\vec{v}_1, \ldots, \vec{v}_k$ so that $\vec{v}_k$ is the vector that can be expressed as a linear combination of the first $k-1$ vectors. Thus we just assume $i = k$. Note that for $i = k$, $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$ is written as $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$.

Similarly, since $\left[\begin{smallmatrix}2\\4\end{smallmatrix}\right] = 2\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] + 4\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, it follows from the Reduction Theorem that

$$U = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}.$$

Finally, since $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ are not scalar multiples of one another, we cannot remove either of them from the spanning set without changing the span. Thus $\vec{x} \in U$ if and only if it satisfies

$$\vec{x} = c_1\begin{bmatrix}1\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\end{bmatrix}$$

for some $c_1, c_2 \in U$ Combining the vectors on the right gives

$$\vec{x} = \begin{bmatrix}c_1\\c_2\end{bmatrix}.$$

Since any $\vec{x} \in \mathbb{R}^2$ has this form, it is clear that $U = \mathbb{R}^2$.

Regarding the last example, the vectors that were chosen to be removed from the spanning set depended on us noticing that some were linear combinations of others. Of course, we could have noticed that $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] = \frac{1}{2}\left[\begin{smallmatrix}2\\4\end{smallmatrix}\right] - 2\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ and concluded that

$$U = \operatorname{Span}\left\{\begin{bmatrix}5\\0\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$$

and then continued from there. Indeed, any of

$$U = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}5\\0\end{bmatrix}, \begin{bmatrix}2\\4\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}5\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}2\\4\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$$

are also correct descriptions of U where the spanning sets cannot be further reduced.

Exercise 56 Use the Reduction Theorem to simplify the spanning set of

$$U = \operatorname{Span}\left\{\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\2\end{bmatrix}, \begin{bmatrix}-5\\-5\\3\end{bmatrix}\right\}$$

by removing any redundant vectors.

[**Hint:** You should end up with a spanning set consisting of two vectors.]

Exercise 57 Refer back to Example 4.2.7.

(a) Find all subsets of

$$S = \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}\right\}$$

containing exactly two vectors so that $\operatorname{Span} U = \operatorname{Span} S$.

(b) Is there a subset of S containing just one vector that spans U?

We have used the the Reduction Theorem as a way of removing dependencies from spanning sets. However, it can also be used to *create* dependencies! This can be useful to show that the spans of two given sets are the same.

Example 4.2.10 Show that

$$\operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\right\}=\operatorname{Span}\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}\right\}.$$

Solution: Using the Reduction Theorem, we have

$$\begin{aligned}
\operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\right\} &= \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix}\right\} && \text{since } \begin{bmatrix}1\\1\end{bmatrix}=1\begin{bmatrix}1\\0\end{bmatrix}+1\begin{bmatrix}0\\1\end{bmatrix}\\
&= \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}\right\} && \text{since } \begin{bmatrix}2\\3\end{bmatrix}=2\begin{bmatrix}1\\0\end{bmatrix}+3\begin{bmatrix}0\\1\end{bmatrix}+0\begin{bmatrix}1\\1\end{bmatrix}\\
&= \operatorname{Span}\left\{\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}\right\} && \text{since } \begin{bmatrix}1\\0\end{bmatrix}=-1\begin{bmatrix}0\\1\end{bmatrix}+1\begin{bmatrix}1\\1\end{bmatrix}+0\begin{bmatrix}2\\3\end{bmatrix}\\
&= \operatorname{Span}\left\{\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}\right\} && \text{since } \begin{bmatrix}0\\1\end{bmatrix}=-2\begin{bmatrix}1\\1\end{bmatrix}+1\begin{bmatrix}2\\3\end{bmatrix}
\end{aligned}$$

Exercise 58 Show that

$$\operatorname{Span}\left\{\begin{bmatrix}1\\1\\2\end{bmatrix},\begin{bmatrix}2\\1\\1\end{bmatrix}\right\}=\operatorname{Span}\left\{\begin{bmatrix}3\\2\\3\end{bmatrix},\begin{bmatrix}-1\\0\\1\end{bmatrix}\right\}.$$

Section 4.2 Problems

4.2.1. Describe the following sets geometrically. Be as descriptive as possible—for example, if you claim something is a line or a plane, give a vector equation for the line or plane.

(a) $S_1 = \operatorname{Span}\left\{\begin{bmatrix}1\\2\end{bmatrix}\right\}$.

(b) $S_2 = \operatorname{Span}\left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\3\\4\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$.

(c) $S_3 = \operatorname{Span}\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}-2\\-2\\0\end{bmatrix}\right\}$.

4.2.2. Use geometry to solve the next two problems.

(a) Let $S_1 = \{\vec{v}_1\} \subseteq \mathbb{R}^2$. Show that $\operatorname{Span} S_1 \neq \mathbb{R}^2$.

(b) Let $S_2 = \{\vec{u}_1, \vec{u}_2\} \subseteq \mathbb{R}^3$. Show that $\operatorname{Span} S_2 \neq \mathbb{R}^3$.

4.2.3. Let

$$S_1 = \left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} \quad \text{and} \quad S_2 = \left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}\right\}.$$

(a) Express each vector in S_1 as a linear combination of the vectors in S_2.

(b) Use part (a) to show that if $\vec{v} \in \operatorname{Span} S_1$, then $\vec{v} \in \operatorname{Span} S_2$, that is, show that $\operatorname{Span} S_1 \subseteq \operatorname{Span} S_2$.

(c) Express each vector in S_2 as a linear combination of the vectors in S_1.

(d) Use part (c) to show that if $\vec{v} \in \operatorname{Span} S_2$, then $\vec{v} \in \operatorname{Span} S_1$, that is, show that $\operatorname{Span} S_2 \subseteq \operatorname{Span} S_1$.

(e) Show that $\operatorname{Span} S_1 = \operatorname{Span} S_2$.

Compare this method of showing the span of two sets are equal to the method presented in Example 4.2.10.

4.2.4. Let $\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1} \in \mathbb{R}^n$, let $S_1 = \{\vec{v}_1, \ldots, \vec{v}_k\}$ and let $S_2 = \{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}\}$. Without using the Reduction Theorem, show that

(a) $\operatorname{Span} S_1 \subseteq \operatorname{Span} S_2$,

(b) if $\operatorname{Span} S_2 \subseteq \operatorname{Span} S_1$, then $\vec{v}_{k+1} \in \operatorname{Span} S_1$,

(c) if $\vec{v}_{k+1} \in \operatorname{Span} S_1$, then $\operatorname{Span} S_2 \subseteq \operatorname{Span} S_1$,

(d) $\operatorname{Span} S_1 = \operatorname{Span} S_2$ if and only if $\vec{v}_{k+1} \in \operatorname{Span} S_1$.

4.2.5. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^n$.

(a) Show that $\operatorname{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \operatorname{Span}\{t_1\vec{v}_1, t_2\vec{v}_2, t_3\vec{v}_3\}$ for any *nonzero* $t_1, t_2, t_3 \in \mathbb{R}$.

(b) Show that $\operatorname{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \operatorname{Span}\{\vec{v}_1, \vec{v}_2 - \vec{v}_1, \vec{v}_3 - \vec{v}_1\}$.

4.3 Linear Dependence and Linear Independence

4.3.1 Definition and Basic Results

In Section 4.2, we discovered that the span of a single vector is not always a line, and that the span of two vectors is not always a plane. More generally, the Reduction Theorem showed that given a set $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$ with $U = \operatorname{Span} S$, we could remove a vector, say $\vec{v}_i$, from S to obtain a smaller set that still spans U if and only if $\vec{v}_i$ could be expressed as a linear combination of the other vectors in S.

Our examples in Section 4.2 were simple enough that we could detect such linear combinations by inspection. However, suppose we are given

$$U = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ -3 \\ 7 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 4 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \\ 8 \\ 2 \end{bmatrix}, \begin{bmatrix} -6 \\ -6 \\ 3 \\ 7 \end{bmatrix} \right\}.$$

It's likely not immediately obvious that

$$\begin{bmatrix} -6 \\ -6 \\ 3 \\ 7 \end{bmatrix} = -\begin{bmatrix} 1 \\ 2 \\ -3 \\ 7 \end{bmatrix} + 2\begin{bmatrix} -2 \\ 1 \\ 4 \\ 8 \end{bmatrix} - \begin{bmatrix} 1 \\ 6 \\ 8 \\ 2 \end{bmatrix},$$

and that we can thus remove the last vector from the spanning set for U. Now imagine being given 500 vectors in $\mathbb{R}^{1000}$ and trying to decide if any one of them is a linear combination of the other 499 vectors. Inspection clearly won't help here, so we need a better way to spot these dependencies among a set of vectors, should they exist. We make a definition here, and will see soon how it can help us identify such dependencies.

Definition 4.3.1

Linear Dependence and Independence

Let $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. We say that S is **linearly dependent** if there exist $c_1, \ldots, c_k \in \mathbb{R}$, *not all zero*, so that

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \vec{0}.$$

We say that S is **linearly independent** if the only solution to

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \vec{0}$$

is $c_1 = \cdots = c_k = 0$, which we call the **trivial solution**.

It is important to understand that by "$c_1, \ldots, c_k$ not all zero", we mean that *at least one of* $c_1, \ldots, c_k$ is nonzero.

Example 4.3.2

Determine whether the set

$$S = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$$

is linearly dependent or linearly independent.

Solution: Let $c_1, c_2 \in \mathbb{R}$ and consider

$$c_1 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Equating entries gives the homogeneous system of linear equations

$$\begin{array}{rcrcl} 2c_1 & - & c_2 & = & 0 \\ 3c_1 & + & 2c_2 & = & 0 \end{array}$$

Reducing the coefficient matrix to row echelon form, we have

$$\begin{bmatrix} 2 & -1 \\ 3 & 2 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 3 & 2 \\ 2 & -1 \end{bmatrix} \xrightarrow{R_1 - R_2} \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix} \xrightarrow[R_2 - 2R_1]{} \begin{bmatrix} 1 & 3 \\ 0 & -7 \end{bmatrix}.$$

We see that there are no free variables, so we get a unique solution. Since the system is homogeneous, the unique solution must be $c_1 = c_2 = 0$, and hence S is linearly independent.

Example 4.3.2 illustrates a useful fact. For a set of two vectors, one can determine linear dependence or linear independence of that set by inspection - if one of vectors is a scalar multiple of the other, then the set is linearly dependent. If neither vector is a scalar multiple of the other, then the set will be linearly independent. This observation *only* works for sets containing two vectors, as the next example illustrates.

Example 4.3.3 Determine whether the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is linearly dependent or linearly independent.

Solution: Let $c_1, c_2, c_3 \in \mathbb{R}$ and consider

$$c_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.5}$$

We obtain

$$\begin{array}{rcrcrcl} c_1 & + & 2c_2 & + & c_3 & = & 0 \\ & & c_2 & + & c_3 & = & 0 \\ -c_1 & & & + & c_3 & = & 0 \end{array}.$$

Carrying the coefficient matrix to row echelon form gives

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \xrightarrow[R_3 + R_1]{} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix} \xrightarrow[R_3 - 2R_2]{} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

from which we see that c_3 is a free variable. We will thus obtain nontrivial solutions, that is, solutions where c_1, c_2, c_3 are not all zero. Hence S is linearly dependent.

Note that in Example 4.3.3, although the set S is linearly dependent, no vector in S is a scalar multiple of any of the other vectors in S.

Exercise 59

Show that the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is linearly independent.

Exercise 60

Determine whether the set

$$S = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \end{bmatrix} \right\}$$

is linearly dependent or independent.

In Example 4.3.2 and Example 4.3.3, we see the appearance of homogeneous systems of linear equations. When checking for linear dependence or linear independence of a set $\{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^n$, we consider the vector equation

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$$

which we see leads to a matrix–vector equation of a homogeneous system of linear equations. We are thus interested in whether or not we have a unique solution (no free variables) or if we have infinitely many solutions (at least one free variable).

Theorem 4.3.4

Let $S = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix}$. Then S is linearly independent if and only if $\operatorname{rank}(A) = k$.

Proof: The set S is linearly independent if and only if the homogeneous system $A\vec{x} = \vec{0}$ has only the trivial solution (as discussed above), which is equivalent to the solution of $A\vec{x} = \vec{0}$ having no free variables. This happens exactly when $\operatorname{rank}(A) = k$ by part(b) of the System–Rank Theorem. □

Example 4.3.5

Consider the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}.$$

Since

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

is already in REF, we see that $\operatorname{rank}(A) = 3$. Thus S is linearly independent by Theorem 4.3.4.

Exercise 61 Determine whether the set

$$S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} \right\}$$

is linearly dependent or linearly independent by computing the rank of the matrix

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}.$$

Example 4.3.6 Consider the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Since

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 2 & 1 \end{bmatrix} \in M_{3\times 4}(\mathbb{R}),$$

we have that $\operatorname{rank}(A) \leq \min\{3, 4\} = 3 < 4$, so S is linearly dependent by Theorem 4.3.4.

Note that in Example 4.3.6, we did not explicitly compute the rank of A, but instead used the fact that A had fewer rows than columns to show that the $\operatorname{rank}(A) < 4$. This, of course, is because $S \subseteq \mathbb{R}^3$ had more than 3 vectors. The following corollary generalizes this observation.

Corollary 4.3.7 Let $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$. If $k > n$, then S is linearly dependent.

Exercise 62 Prove Corollary 4.3.7.

It follows from Corollary 4.3.7 that if $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$ is linearly independent, then $k \leq n$, that is, S can have at most n vectors if it is to be linearly independent. However, S having $k \leq n$ vectors does not guarantee that S is linearly independent.

Exercise 63 Give an example of set $S \subseteq \mathbb{R}^4$ containing 3 vectors that is linearly dependent.

Returning to Example 4.3.3, note that we did not solve the resulting homogeneous system of linear equations since we only wanted to know if it had any non-trivial solutions. However, once we have discovered that there are non-trivial solutions – that is, once we have determined that S is linearly *dependent* – we can then use these non-trivial solutions to find dependencies in S. This will allow us to get a better understanding of $\operatorname{Span} S$ by removing any redundant vectors from the spanning set S. We illustrate this in the next example.

Example 4.3.8 Let

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

In Example 4.3.3, we saw that the equation

$$c_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.6}$$

led us to a homogeneous system of linear equations, whose coefficient matrix we can row reduce to

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \xrightarrow[R_3+R_1]{} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_1-2R_2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

From this, we see that $c_1 = t$, $c_2 = -t$ and $c_3 = t$ for any $t \in \mathbb{R}$. Substituting these values into (4.6) gives

$$t \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} - t \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Choosing any solution with $t \neq 0$ will allow us to detect dependencies between the vectors in S. For instance, if we choose $t = 1$, we get

$$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.7}$$

We may rearrange this as

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

which allows us to use the Reduction Theorem to conclude that

$$\operatorname{Span} S = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} = \operatorname{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Note that the new spanning set $\left\{ \left[\begin{smallmatrix} 2 \\ 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right] \right\}$ is linearly independent, since the two vectors it contains are not scalar multiples of each other, so it cannot be further reduced. We conclude $\operatorname{Span} S$ is a plane in $\mathbb{R}^3$ through the origin.

Be aware that we had some freedom in how we rearranged (4.7). We could have solved for any vector on the left hand side of (4.7) in terms of the other two to alternatively arrive at

$$\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Longrightarrow \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

or

$$\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\1\\0\end{bmatrix} - \begin{bmatrix}1\\0\\-1\end{bmatrix} \Longrightarrow \operatorname{Span}\left\{\begin{bmatrix}1\\0\\-1\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\-1\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}\right\}.$$

In either case, we arrive at spanning sets for Span S that are linearly independent since they contain just two vectors that are not scalar multiples of one another.

In Example 4.3.8, we observed that S is linearly dependent, and that we are able to remove any one of the three vectors from S in order to obtain a linear independent set of two vectors with the same span as S. The next example shows that we can't always arbitrarily remove a vector from a linearly dependent set.

Example 4.3.9 Consider the set

$$S = \left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\-1\end{bmatrix}\right\}.$$

Let $c_1, c_2, c_3 \in \mathbb{R}$ and consider

$$c_1\begin{bmatrix}1\\0\\1\end{bmatrix} + c_2\begin{bmatrix}0\\1\\0\end{bmatrix} + c_3\begin{bmatrix}-1\\0\\-1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}. \tag{4.8}$$

Equating entries gives rise to a homogeneous system of linear equations whose coefficient matrix we carry to row echelon form to obtain

$$\begin{bmatrix}1&0&-1\\0&1&0\\1&0&-1\end{bmatrix} \underset{R_3-R_1}{\longrightarrow} \begin{bmatrix}1&0&-1\\0&1&0\\0&0&0\end{bmatrix}.$$

We see that the rank of the coefficient matrix is $2 < 3$, so S is linearly dependent. We solve the system to obtain $c_1 = t$, $c_2 = 0$ and $c_3 = t$ for any $t \in \mathbb{R}$. Substituting these values into (4.8) gives

$$t\begin{bmatrix}1\\0\\1\end{bmatrix} + 0\begin{bmatrix}0\\1\\0\end{bmatrix} + t\begin{bmatrix}-1\\0\\-1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Choosing $t = 1$ (or any solution with $t \neq 0$) gives

$$\begin{bmatrix}1\\0\\1\end{bmatrix} + 0\begin{bmatrix}0\\1\\0\end{bmatrix} + \begin{bmatrix}-1\\0\\-1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}. \tag{4.9}$$

Notice that in (4.9), we can only solve for $\begin{bmatrix}1\\0\\1\end{bmatrix}$ or $\begin{bmatrix}-1\\0\\-1\end{bmatrix}$. Doing so and applying the Reduction Theorem gives either

$$\begin{bmatrix}1\\0\\1\end{bmatrix} = 0\begin{bmatrix}0\\1\\0\end{bmatrix} - \begin{bmatrix}-1\\0\\-1\end{bmatrix} \Longrightarrow \operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\-1\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\-1\end{bmatrix}\right\}$$

or

$$\begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} = 0\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \Longrightarrow \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}.$$

In either case, we cannot reduce the spanning set any further since neither of the two vectors remaining in the spanning is are a scalar multiple of the other. This shows that Span S is a plane through the origin in $\mathbb{R}^3$.

Note that in Example 4.3.9, we are unable to isolate for $\left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right]$ in (4.9) due to the zero coefficient. As a consequence, we cannot remove $\left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right]$ from S. Indeed, without $\left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right]$, we are left with

$$\operatorname{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$$

which is a line through the origin, not a plane.

Example 4.3.8 and Example 4.3.9 seem to indicate that if a set is linearly dependent, then *at least* one of the vectors in that set is a linear combination of the other vectors. The following theorem shows that this is indeed always the case.

Theorem 4.3.10 **(Dependency Theorem)**

A set of vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ in $\mathbb{R}^n$ is linearly dependent if and only if

$$\vec{v}_i \in \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_k\}$$

for some $i = 1, \dots, k$.

Proof: Assume first that the set $\{\vec{v}_1, \dots, \vec{v}_k\}$ in $\mathbb{R}^n$ is linearly dependent. Then there exist $c_1, \dots, c_k \in \mathbb{R}$, not all zero, such that

$$c_1\vec{v}_1 + \cdots + c_{i-1}\vec{v}_{i-1} + c_i\vec{v}_i + c_{i+1}\vec{v}_{i+1} + \cdots + c_k\vec{v}_k = \vec{0}.$$

Without loss of generality, assume that $c_i \neq 0$. Then we may isolate for $\vec{v}_i$ on one side of the equation:

$$\vec{v}_i = -\frac{c_1}{c_i}\vec{v}_1 - \cdots - \frac{c_{i-1}}{c_i}\vec{v}_{i-1} - \frac{c_{i+1}}{c_i}\vec{v}_{i+1} - \cdots - \frac{c_k}{c_i}\vec{v}_k$$

which shows that $\vec{v}_i \in \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_k\}$. To prove the other implication, we assume that $\vec{v}_i \in \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_{i-1}, \vec{v}_{i+1}, \dots, \vec{v}_k\}$ for some $i = 1, \dots, k$. Then there exist $d_1, \dots, d_{i-1}, d_{i+1}, \dots, d_k \in \mathbb{R}$ such that

$$\vec{v}_i = d_1\vec{v}_1 + \cdots + d_{i-1}\vec{v}_{i-1} + d_{i+1}\vec{v}_{i+1} + \cdots + d_k\vec{v}_k$$

and rearranging gives

$$d_1\vec{v}_1 + \cdots + d_{i-1}\vec{v}_{i-1} - 1\vec{v}_i + d_{i+1}\vec{v}_{i+1} + \cdots + d_k\vec{v}_k = \vec{0}$$

which shows that $\{\vec{v}_1, \dots, \vec{v}_k\}$ is linearly dependent. $\square$

Exercise 64 Let

$$S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 6 \\ 6 \end{bmatrix} \right\}$$

and let $U = \operatorname{Span} S$. Find all subsets of S that are linearly independent spanning sets for U, and describe U geometrically.

We can summarize our finding as follows. Consider $U = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$. If $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly *dependent*, then the Dependency Theorem guarantees that there must be some vector in S (say $\vec{v}_i$) that is a linear combination of the others. The Reduction Theorem then tells us we can remove $\vec{v}_i$ from S without affecting the span:

$$U = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}.$$

4.3.2 Removing Dependencies

The spanning sets presented in Example 4.3.8 and Example 4.3.9 (along with Exercise 64) were quite simple - they were each linearly dependent sets containing 3 vectors, and after removing an appropriately chosen vector, each reduced to a set of 2 vectors that was clearly linearly independent. Example 4.2.9 gives an instance where were needed to remove two vectors from the given spanning set before arriving at a linearly independent spanning set.[3].

In general, things become more complicated when S contains many vectors and we cannot detect dependencies by inspection - our current method of repeatedly checking for linear dependence and removing an appropriate vector until an independent set is obtained can become increasingly tedious as we increase the number of vectors in S. The solution to the next example will lead to a more efficient method to handle this problem

Example 4.3.11 Let

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix} \right\}$$

and let $U = \operatorname{Span} S$. Find a linearly independent subset of S that is also a spanning set for U.

Solution: For $c_1, c_2, c_3, c_4 \in \mathbb{R}$, consider

$$c_1 \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix} + c_4 \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.10}$$

Carrying the coefficient matrix of this homogeneous system of linear equations to reduced row echelon form gives

$$\begin{bmatrix} 1 & 1 & 3 & 4 \\ 1 & -1 & -1 & 0 \\ -1 & 3 & 5 & 4 \\ 2 & 1 & 4 & 6 \end{bmatrix} \underset{\substack{R_2 - R_1 \\ R_3 + R_1 \\ R_4 - 2R_1}}{\longrightarrow} \begin{bmatrix} 1 & 1 & 3 & 4 \\ 0 & -2 & -4 & -4 \\ 0 & 4 & 8 & 8 \\ 0 & -1 & -2 & -2 \end{bmatrix} \underset{-\frac{1}{2}R_2}{\longrightarrow} \begin{bmatrix} 1 & 1 & 3 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 4 & 8 & 8 \\ 0 & -1 & -2 & -2 \end{bmatrix} \underset{\substack{R_3 - 4R_2 \\ R_4 + R_2}}{\overset{R_1 - R_2}{\longrightarrow}} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

[3]Note that Example 4.2.9 occured *before* the definition of linear dependence/independence.

Since the rank of the coefficient matrix is $2 < 4$, we have that S is linearly dependent by Theorem 4.3.4. It follows from the Dependency Theorem that at least one vector from S can be removed while still having a spanning set for U. To find which vectors we can remove, we solve the homogeneous system, which gives

$$c_1 = -s - 2t, \quad c_2 = -2s - 2t, \quad c_3 = s \quad \text{and} \quad c_4 = t$$

for any $s, t \in \mathbb{R}$. We can thus rewrite (4.10) as

$$(-s-2t)\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix} + (-2s-2t)\begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix} + s\begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix} + t\begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.11}$$

Taking $s = 1$ and $t = 0$, we may solve for $\left[\begin{smallmatrix} 3 \\ -1 \\ 5 \\ 4 \end{smallmatrix}\right]$ in (4.11) to obtain

$$\begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \tag{4.12}$$

while taking $s = 1$ and $t = 0$, we may solve for $\left[\begin{smallmatrix} 4 \\ 0 \\ 4 \\ 6 \end{smallmatrix}\right]$ in (4.11) which gives

$$\begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}. \tag{4.13}$$

We see from (4.12) that

$$\begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix} \in \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}\right\} \subseteq \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix}\right\},$$

so we apply the Reduction Theorem to obtain

$$U = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix}\right\}$$

From (4.13), we have

$$\begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix} \in \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}\right\},$$

so the Reduction Theorem gives that

$$U = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}\right\}.$$

Since neither of the vectors in the set

$$S' = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix} \right\}$$

are scalar multiples of the other, S' is a linearly independent subset of S that spans U.

Observe that the vectors we removed from S in Example 4.3.11 depended on our choices for the parameters s and t as well as which vectors in (4.10) we solved for based on these parameters. For this example, it can be shown *any* subset of S containing exactly two vectors will be linearly independent and span U.

However, the solution to Example 4.3.11 can be simplified so that we are not required to explicitly deal with parameters. Let A be the coefficient matrix of the homogeneous system in Example 4.3.11 and let R be the reduced row echelon form of A, that is, let

$$A = \begin{bmatrix} 1 & 1 & 3 & 4 \\ 1 & -1 & -1 & 0 \\ -1 & 3 & 5 & 4 \\ 2 & 1 & 4 & 6 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that the columns of A are exactly the vectors in S, in exactly the same order. While it is difficult to detect any dependencies among that columns of A by mere inspection, it very easy to determine dependencies among the columns of R without any work: the first two columns of R (which have leading ones) are the first two standard basis vectors $\vec{e}_1$ and $\vec{e}_2$ of $\mathbb{R}^4$ and the last two columns of R (which do not have leading ones) are easily seen to be linear combination of the first two columns of R:

$$\begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}. \tag{4.14}$$

Our work in Example 4.3.11 shows the last two columns of A can be expressed as linear combination of the first two columns of A using the *exact* same coefficients:

$$\begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}. \tag{4.15}$$

From (4.14), we see that applying the Reduction Theorem twice gives that

$$\operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix} \right\} = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\},$$

while in Example 4.3.11 we saw that (4.15) prompted two applications of the Reduction Theorem which lead to

$$\operatorname{Span}\left\{\begin{bmatrix}1\\1\\-1\\2\end{bmatrix},\begin{bmatrix}1\\-1\\3\\1\end{bmatrix},\begin{bmatrix}3\\-1\\5\\4\end{bmatrix},\begin{bmatrix}4\\0\\4\\6\end{bmatrix}\right\}=\operatorname{Span}\left\{\begin{bmatrix}1\\1\\-1\\2\end{bmatrix},\begin{bmatrix}1\\-1\\3\\1\end{bmatrix}\right\}.$$

Finally, both of the sets

$$\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right\}\quad\text{and}\quad\left\{\begin{bmatrix}1\\1\\-1\\2\end{bmatrix},\begin{bmatrix}1\\-1\\3\\1\end{bmatrix}\right\}$$

are linearly independent - the first because each vector has a 1 where the other has a 0, and the second as verified in Example 4.3.11. Notice that the vectors in the second set are simply the columns of A that correspond to those columns of R that have leading entries.

Given a set $S=\{\vec{v}_1,\dots,\vec{v}_k\}\subseteq\mathbb{R}^n$, our above discussion supports the idea that we can use the reduced row echelon form of $A=\begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_k\end{bmatrix}$ to extract a linearly independent subset B of S with $\operatorname{Span} S'=\operatorname{Span} S$. The next theorem shows that this is indeed the case.

Theorem 4.3.12 **(Extraction Theorem)**

Given $S=\{\vec{v}_1,\dots,\vec{v}_k\}\subseteq\mathbb{R}^n$, let $A=\begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_k\end{bmatrix}$ and let R be any row echelon form of A. The subset S' of S defined by

$$\vec{v}_i\in B \text{ if and only if the } i\text{th column of } R \text{ contains a leading entry}$$

is linearly independent with $\operatorname{Span} S'=\operatorname{Span} S$.

Proof: For $c_1,\dots,c_k\in\mathbb{R}$, consider

$$c_1\vec{v}_1+\cdots+c_k\vec{v}_k=\vec{0}.$$

We express this homogeneous system of linear equations by the matrix–equation $A\vec{x}=\vec{0}$. Let $R=\begin{bmatrix}\vec{w}_1 & \cdots & \vec{w}_k\end{bmatrix}$ be the *reduced* row echelon form of A. Then the systems $A\vec{x}=\vec{0}$ and $R\vec{x}=\vec{0}$ are equivalent, that is, they have the exact same set of solutions. Hence

$$c_1\vec{v}_1+\cdots+c_k\vec{v}_k=\vec{0}\quad\text{if and only if}\quad c_1\vec{w}_1+\cdots+c_k\vec{w}_k=\vec{0},\tag{4.16}$$

and it follows that the exact same dependencies exist among the columns of A and the columns of R. Let $r=\operatorname{rank}(A)\le k$. Then there are r columns of R that contain leading ones, say $\vec{w}_{i_1},\dots,\vec{w}_{i_r}$, where

$$1\le i_1<i_2<\cdots<i_r\le k.$$

Since R is in reduced row echelon form, we have that

$$\vec{w}_{i_1}=\vec{e}_1,\quad\vec{w}_{i_2}=\vec{e}_2,\quad\dots,\quad\vec{w}_{i_r}=\vec{e}_r,$$

while the remaining $k-r$ columns of R can be expressed as linear combinations of $\vec{w}_{i_1}, \vec{w}_{i_2}, \ldots, \vec{w}_{i_r}$. After $k-r$ applications of the Reduction Theorem, we are thus left with

$$\operatorname{Span}\{\vec{w}_1, \ldots, \vec{w}_k\} = \operatorname{Span}\{\vec{w}_{i_1}, \ldots, \vec{w}_{i_r}\},$$

and it follows from (4.16) that we may remove the same $k-r$ columns from S to obtain

$$\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_{i_1}, \ldots, \vec{v}_{i_r}\},$$

that is, $\operatorname{Span} S' = \operatorname{Span} S$ where $S' = \{\vec{v}_{i_1}, \ldots, \vec{v}_{i_r}\}$ is clearly a subset of S.

It remains to show that S' is linearly independent. Since the reduced row echelon form of $\begin{bmatrix} \vec{v}_{i_1} & \cdots & \vec{v}_{i_r} \end{bmatrix}$ is $\begin{bmatrix} \vec{w}_{i_1}, \ldots, \vec{w}_{i_r} \end{bmatrix} = \begin{bmatrix} \vec{e}_1 & \cdots & \vec{e}_r \end{bmatrix}$ and $\operatorname{rank}\left(\begin{bmatrix} \vec{e}_1 & \cdots & \vec{e}_r \end{bmatrix}\right) = r$, we have that S' is linear independent by Theorem 4.3.4.

Finally, we note that we need only carry A to row echelon form, since any row echelon form of A will have leading entries in exactly the same columns that R has leading ones in. $\square$

Despite a somewhat complicated proof, the Extraction Theorem is extremely easy to use. Given a set $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$, simply construct $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix}$ and carry A to any row echelon form. Note the columns of the row echelon form that contain leading entries and construct a set S' containing only the corresponding columns of A. The resulting set S' will be linearly independent with $\operatorname{Span} S' = \operatorname{Span} S$.

Let's look at an example.

Example 4.3.13 Let

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

and let $U = \operatorname{Span} S$. Determine a linearly independent subset S' of S that also spans U.

Solution: We carry

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

to row echelon form:

$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 1 & 0 & -3 & -1 \\ 2 & 1 & 2 & 1 \\ -1 & -1 & -1 & 0 \end{bmatrix} \underset{\substack{R_2-R_1 \\ R_3-2R_1 \\ R_4+R_1}}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & -1 & -2 & 0 \end{bmatrix} \underset{R_2 \leftrightarrow R_4}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \underset{R_3+R_2}{\longrightarrow}$$

$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \underset{R_4+R_3}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

As there are leading entries in the first, second and last columns of an REF of A, the Extraction Theorem states that we may take the first, second and last columns of A as our linearly independent spanning set for U. Thus,

$$S' = \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a linearly independent subset of S with $U = \operatorname{Span} S'$.

In Example 4.3.13, let

$$R = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Although the same dependencies exist among the columns of A and the columns of R, the columns of R generally do not span the same set as the columns of A. Thus, after determining which columns of R have leading entries, we must choose the corresponding columns from A, and not R. Indeed,

$$\operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\} \neq \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

as the fourth entry of any vector in the former set must be 0 while this is not so for the latter set. Thus it would be incorrect to take

$$S' = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

in Example 4.3.13.

Additionally, note that when constructing the matrix A in Example 4.3.13, the way in which we order of the columns can impact the linearly independent set S' that we wish to construct. For example, if we had interchanged the first two vectors in S, the we would constructed the matrix

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

Then we would find that a row echelon form of A is

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

and the Extraction Theorem would lead us to choose

$$S' = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 2 \\ -1 \end{bmatrix} \right\}$$

instead. Note that this is also a correct solution for Example 4.3.13, but that it is not the same linearly independent spanning set as what we derived above.

Exercise 65 Let

$$S = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ -2 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -1 \\ 8 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

and let $U = \operatorname{Span} S$. Determine a linearly independent subset S' of S that also spans U.

4.3.3 Further Examples

We conclude with a few more examples involving linear dependence and linear independence. The first shows that if a set of k vectors contains the zero vector, then the set is linearly dependent.

Example 4.3.14 Consider the set $\{\vec{v}_1, \ldots, \vec{v}_{k-1}, \vec{0}\}$ of vectors in $\mathbb{R}^n$. We will show that this set is linearly dependent in two ways.

First, we observe that

$$0\vec{v}_1 + \cdots + 0\vec{v}_{k-1} + (1)\vec{0} = \vec{0}$$

which, by Definition 4.3.1 shows that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}, \vec{0}\}$ is linearly dependent.

Second, we note that since

$$\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_{k-1},$$

we have that

$$\vec{0} \in \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$$

so $\{\vec{v}_1, \ldots, \vec{v}_{k-1}, \vec{0}\}$ is linearly dependent by the Dependency Theorem

It's useful to compare the solutions presented in Example 4.3.14 to your solutions for Exercise 47 and Exercise 49.

The next example shows how we can use the linear independence of one set to show the linear independence of another set.

Example 4.3.15 Let $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^n$ be such that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent. Prove that the set

$$\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$$

is linearly independent.

Proof: We must prove that the set $\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$ is linearly independent. To do so, we consider the vector equation

$$c_1\vec{v}_1 + c_2(\vec{v}_1 + \vec{v}_2) + c_3(\vec{v}_1 + \vec{v}_2 + \vec{v}_3) = \vec{0}, \quad c_1, c_2, c_3 \in \mathbb{R}.$$

Rearranging this equation gives

$$(c_1 + c_2 + c_3)\vec{v}_1 + (c_2 + c_3)\vec{v}_2 + c_3\vec{v}_3 = \vec{0}.$$

Since $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent, we must have that

$$\begin{array}{rcrcrcl} c_1 & + & c_2 & + & c_3 & = & 0 \\ & & c_2 & + & c_3 & = & 0 \\ & & & & c_3 & = & 0 \end{array}$$

We see that $c_3 = 0$ and it follows that $c_2 = 0$ and then that $c_1 = 0$. Hence we have only the trivial solution, so our set $\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$ is linearly independent.

Finally, we consider nonempty subsets of a linearly independent set.

Example 4.3.16 Let $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be a linearly independent set of vectors in $\mathbb{R}^n$ with $k \geq 2$. Prove that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is linearly independent.

Proof: It is given that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent. Suppose for a contradiction that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is linearly dependent. Then there exist $c_1, \ldots, c_{k-1}$, not all zero, such that

$$c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} = \vec{0}.$$

But then adding $0\vec{v}_k$ to both sides gives

$$c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} + 0\vec{v}_k = \vec{0}$$

which shows that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly dependent, since not all of $c_1, \ldots, c_{k-1}$ are zero. But this is a contradiction since we were given that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent. Hence, our supposition that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is linearly dependent was incorrect. This leaves only that $\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is linearly independent, as required.

In the solution of Example 4.3.16, we used a proof technique known as *Proof by Contradiction*. When using proof by contradiction, you are essentially proving a statement is true by proving that it cannot be false. We are told that the set $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent and asked to show that under this assumption, the set $S' = \{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$ is also linearly independent. The set S' must be either linearly independent or linearly dependent, but not both. So instead of proving that S' is linearly independent directly, we suppose that S' is linearly dependent. From that supposition, we argue until we arrived at S being linearly dependent, which is impossible since we are given that S is linearly independent as part of our hypothesis. S being linearly dependent is thus a *contradiction*. Since this contradiction was derived from our supposition that S' is linearly dependent, the supposition that S' is linearly dependent is incorrect. Since S' is not linearly dependent, it must be linearly independent (which is what we were asked to prove).

It follows from the last example that every nonempty subset of a linearly independent set is also linearly independent. Of course, we should consider the empty set, $\varnothing$, since it is a subset of every set. As the empty set contains no vectors, we cannot exhibit vectors from the empty set that form a linearly dependent set. Thus, the empty set is (vacuously) linearly independent. Thus, we can now say that given any linearly independent set S, *every* subset of S is linearly independent as well.

Section 4.3 Problems

4.3.1. For each of the following, determine if S is linearly dependent or linearly independent. If S is linearly dependent, express each vector in S as a linear combination of the other vectors in S whenever possible.

(a) $S = \left\{ \begin{bmatrix} 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 7 \\ 2 \end{bmatrix} \right\}$.

(b) $S = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 15 \\ 21 \end{bmatrix} \right\}$.

(c) $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 10 \end{bmatrix} \right\}$.

(d) $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ -2 \end{bmatrix}, \begin{bmatrix} 10 \\ 19 \\ -6 \end{bmatrix} \right\}$.

(e) $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.

4.3.2. For each of the given sets S, find a linearly independent subset S' of S with $\operatorname{Span} S' = \operatorname{Span} S$.

(a) $S = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$.

(b) $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 10 \end{bmatrix} \right\}$.

(c) $S = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 5 \\ -1 \\ -7 \end{bmatrix} \right\}$.

(d) $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\}$.

(e) $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 4 \end{bmatrix} \right\}$.

4.3.3. Let $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ be a linearly independent set of vectors in $\mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Define

$$\vec{v}_1 = \vec{x}_1 - \alpha\vec{x}_3, \quad \vec{v}_2 = \vec{x}_2 - \alpha\vec{x}_1 \quad \text{and} \quad \vec{v}_3 = \vec{x}_3 - \alpha\vec{x}_2.$$

For which values of $\alpha \in \mathbb{R}$ is the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ linearly dependent?

4.3.4. Let $\{\vec{x}, \vec{y}, \vec{z}\} \subseteq \mathbb{R}^n$ be a linearly dependent set. Prove that if $\vec{z} \notin \operatorname{Span}\{\vec{x}, \vec{y}\}$, then either $\vec{x}$ is a scalar multiple of $\vec{y}$, or $\vec{y}$ is a scalar multiple of $\vec{x}$.

4.3.5. Let $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4 \in \mathbb{R}^n$ be such that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is linearly independent. For each of the following sets, determine whether they are linearly independent or linearly dependent.

(a) $\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_3, \vec{v}_1 + \vec{v}_4\}$.
(b) $\{\vec{v}_1 - \vec{v}_2, \vec{v}_2 - \vec{v}_3, \vec{v}_3 - \vec{v}_4, \vec{v}_4 - \vec{v}_1\}$.
(c) $\{A\vec{v}_1, A\vec{v}_2, A\vec{v}_3, A\vec{v}_4\}$, where $A \in M_{n\times n}(\mathbb{R})$ is an invertible matrix.

4.3.6. Prove or disprove the following statements.

(a) If the sets $\{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$ and $\{\vec{w}_1, \ldots, \vec{w}_\ell\} \subseteq \mathbb{R}^n$ are linearly independent, then $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{w}_1, \ldots, \vec{w}_\ell\}$ is linearly independent.
(b) If the sets $\{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \mathbb{R}^n$ and $\{\vec{w}_1, \ldots, \vec{w}_\ell\} \subseteq \mathbb{R}^n$ are linearly independent and $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} \cap \operatorname{Span}\{\vec{w}_1, \ldots, \vec{w}_\ell\} = \{\vec{0}\}$, then $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{w}_1, \ldots, \vec{w}_\ell\}$ is linearly independent.

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

We have seen that linear combinations have played a significant role throughout the course, particularly so in Section 4.1 through Section 4.3. Recall from the Fundamental Properties of Vector Algebra that $\mathbb{R}^n$ is closed under vector addition and scalar multiplication and that from these two facts, it followed that $\mathbb{R}^n$ is closed under linear combinations. This means that given vectors $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$ and scalars $c_1, \ldots, c_k \in \mathbb{R}^n$, the vector $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$ is also a vector in $\mathbb{R}^n$. We will now be interested in those subsets of $\mathbb{R}^n$ that are also closed under linear combinations.

Definition 4.4.1
Subspace

A subset U of $\mathbb{R}^n$ is a **subspace** of $\mathbb{R}^n$ if the following properties are all satisfied:

S1. $\vec{0}_{\mathbb{R}^n} \in U$ — U contains the zero vector of $\mathbb{R}^n$

S2. if $\vec{x}, \vec{y} \in U$, then $\vec{x} + \vec{y} \in U$ — U is closed under vector addition

S3. if $\vec{x} \in U$ and $c \in \mathbb{R}$, then $c\vec{x} \in U$ — U is closed under scalar multiplication

The condition S1 guarantees that $U \subseteq \mathbb{R}^n$ is nonempty, and we normally write $\vec{0}$ instead of $\vec{0}_{\mathbb{R}^n}$ as it is clear we are talking about the zero vector of $\mathbb{R}^n$. If U then satisfies S2 and S3, then it will be closed under linear combinations, that is, if $\vec{v}_1, \ldots, \vec{v}_k \in U$ and $c_1, \ldots, c_k \in \mathbb{R}$, then the vector $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \in U$.

Example 4.4.2

We have that $\mathbb{R}^n$ is itself a subspace of $\mathbb{R}^n$. To see this, note that $\vec{0} \in \mathbb{R}^n$ so S1 holds. That S2 and S3 hold follows immediately from the Fundamental Properties of Vector Algebra: S 2 is simply V1 and S3 is just V4.

Exercise 66

Show that $U = \{\vec{0}\}$ is a subspace of $\mathbb{R}^n$. (This is called the *trivial subspace* of $\mathbb{R}^n$.)

Example 4.4.3

The set

$$U = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

is not a subspace of $\mathbb{R}^2$ since $\vec{0} \notin U$, that is, S1 fails.

Example 4.4.3 demonstrates that it's easy to show a subset of $\mathbb{R}^n$ is not a subspace of $\mathbb{R}^n$ if $\vec{0} \notin U$. We also note that since $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] + \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right] \notin U$, U is not closed under vector addition, and since $2\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right] \notin U$, U is not closed under scalar multiplication. Thus S1, S3 and S2 all fail. It is enough to show that just one of S1, S3 and S2 fail to conclude that U is not a subspace of $\mathbb{R}^n$.

Example 4.4.4

Show that

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; x_1 - x_2 + 2x_3 = 0 \right\}$$

is a subspace of $\mathbb{R}^3$.

Solution: We verify S1, S2 and S3.

S1: We must show that $\vec{0} = \left[\begin{smallmatrix}0\\0\\0\end{smallmatrix}\right] \in U$. But since $0 - 0 + 2(0) = 0$, we see easily that $\vec{0} \in U$. Thus S1 holds.

S2: Let

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

be vectors in U. Then they satisfy the condition to belong to U, namely

$$y_1 - y_2 + 2y_3 = 0 \quad \text{and} \quad z_1 - z_2 + 2z_3 = 0 \tag{4.17}$$

We must show that

$$\vec{y} + \vec{z} = \begin{bmatrix} y_1 + z_1 \\ y_2 + z_2 \\ y_3 + z_3 \end{bmatrix} \in U$$

by showing that $(y_1 + z_1) - (y_2 + z_2) + 2(y_3 + z_3) = 0$. We have

$$\begin{aligned} (y_1 + z_2) - (y_2 + z_2) + 2(y_3 + z_3) &= (y_1 - y_2 + 2y_3) + (z_1 - z_2 + 2z_3) \\ &= 0 + 0 && \text{by (4.17)} \\ &= 0, \end{aligned}$$

so $\vec{y} + \vec{z} \in U$ and S2 holds.

S3: Let $c \in \mathbb{R}$ and $\vec{y}$ be as above. We must show that

$$c\vec{y} = \begin{bmatrix} cy_1 \\ cy_2 \\ cy_3 \end{bmatrix} \in U$$

by showing that $cy_1 - cy_2 + 2cy_3 = 0$. We have

$$\begin{aligned} (cy_1) - (cy_2) + 2(cy_3) &= c(y_1 - y_2 + 2y_3) \\ &= c(0) && \text{by (4.17)} \\ &= 0, \end{aligned}$$

so $c\vec{y} \in U$ and S3 holds.

Thus U is a subspace of $\mathbb{R}^3$.

Exercise 67 Show that

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; x_1 - x_2 + 2x_3 = 4 \right\}$$

is not a subspace of $\mathbb{R}^3$.

In practice, we don't normally explicitly list S1, S2 and S3, and we don't normally state what we are going to do before we do it (although these are not bad habits to maintain as you begin learning the material!).

Example 4.4.5 Show that

$$U = \left\{ t \begin{bmatrix} 1 \\ 3 \end{bmatrix} \middle| \, t \in \mathbb{R} \right\}$$

is a subspace of $\mathbb{R}^2$.

Solution: Taking $t = 0$ gives $\vec{0} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] \in U$. Now let $\vec{x}, \vec{y} \in U$. Then there exist $t_1, t_2 \in \mathbb{R}$ such that

$$\vec{x} = t_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{y} = t_2 \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

It follows that

$$\vec{x} + \vec{y} = t_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ 3 \end{bmatrix} = (t_1 + t_2) \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

where $t_1 + t_2 \in \mathbb{R}$. Thus $\vec{x} + \vec{y} \in U$, which shows that U is closed under vector addition. For any $c \in \mathbb{R}$,

$$c\vec{x} = c\left(t_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right) = (ct_1) \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

where $ct_1 \in \mathbb{R}$. Thus $c\vec{x} \in U$, which shows that U is closed under scalar multiplication. Hence U is a subspace of $\mathbb{R}^2$.

Example 4.4.6 Show that

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \,\middle|\, x_1 + x_2 = 0 \text{ and } x_2 - x_3 = 0 \right\}$$

is a subspace of $\mathbb{R}^3$.

Solution: Since $0 + 0 = 0$ and $0 - 0 = 0$, $\vec{0} \in U$. Now let $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]$ and $\vec{y} = \left[\begin{smallmatrix} y_1 \\ y_2 \\ y_3 \end{smallmatrix}\right]$ be two vectors in U. Then $x_1 + x_2 = 0 = x_2 - x_3$ and $y_1 + y_2 = 0 = y_2 - y_3 = 0$. For $\vec{x} + \vec{y} = \left[\begin{smallmatrix} x_1+y_1 \\ x_2+y_2 \\ x_3+y_3 \end{smallmatrix}\right]$, we have

$$(x_1 + y_1) + (x_2 + y_2) = (x_1 + x_2) + (y_1 + y_2) = 0 + 0 = 0$$

and

$$(x_2 + y_2) - (x_3 + y_3) = (x_2 - x_3) + (y_2 - y_3) = 0 + 0 = 0$$

so $\vec{x} + \vec{y} \in U$. For $c\vec{x} = \left[\begin{smallmatrix} cx_1 \\ cx_2 \\ cx_3 \end{smallmatrix}\right]$ with $c \in \mathbb{R}$, we have

$$cx_1 + cx_2 = c(x_1 + x_2) = c(0) = 0$$

and

$$cx_2 - cx_3 = c(x_2 - x_3) = c(0) = 0$$

so $c\vec{x} \in U$. Hence U is a subspace of $\mathbb{R}^3$.

The next theorem shows that given a set $\{\vec{v}_1, \ldots, \vec{v}_k\}$ of vectors in $\mathbb{R}^n$, the span of that set will always be a subspace of $\mathbb{R}^n$.

Theorem 4.4.7 Let $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$. Then $U = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a subspace of $\mathbb{R}^n$.

Proof: Clearly we have $\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_k \in U$. Now let $\vec{x}, \vec{y} \in U$. Then there exist $c_1, \ldots, c_k, d_1, \ldots, d_k \in \mathbb{R}$ such that

$$\vec{x} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \quad \text{and} \quad \vec{y} = d_1\vec{v}_1 + \cdots + d_k\vec{v}_k.$$

Then

$$\vec{x} + \vec{y} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k + d_1\vec{v}_1 + \cdots + d_k\vec{v}_k = (c_1 + d_1)\vec{v}_1 + \cdots + (c_k + d_k)\vec{v}_k$$

and so $\vec{x} + \vec{y} \in U$ as it is a linear combination of $\vec{v}_1, \ldots, \vec{v}_k$. For any $c \in \mathbb{R}$,

$$c\vec{x} = c(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) = (cc_1)\vec{v}_1 + \cdots + (cc_k)\vec{v}_k$$

from which we see that $c\vec{x} \in U$ as it is also a linear combination of $\vec{v}_1, \ldots, \vec{v}_k$. Thus, U is a subspace of $\mathbb{R}^n$. $\square$

Theorem 4.4.7 shows that we can always generate a subspace of $\mathbb{R}^n$ by taking the span of a finite set of vectors from $\mathbb{R}^n$. In fact, the next theorem shows that every subspace U of $\mathbb{R}^n$ can be expressed in this way.

Theorem 4.4.8 Let U be a subspace of $\mathbb{R}^n$. Then there are vectors $\vec{v}_1, \ldots, \vec{v}_k \in U$ so that

$$U = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}.$$

Proof: If U is the trivial subspace, that is, if $U = \{\vec{0}\}$, then $U = \operatorname{Span}\{\vec{0}\}$ and we are done. Assume then that U is not the trivial subspace.

Picking a nonzero vector $\vec{v}_1 \in U$, we see that $\{\vec{v}_1\}$ is linearly independent and that $\operatorname{Span}\{\vec{v}_1\} \subseteq U$ since U is closed under scalar multiplication. If $U = \operatorname{Span}\{\vec{v}_1\}$, then we are done.

Otherwise, $U \neq \operatorname{Span}\{\vec{v}_1\}$ so there exists a $\vec{v}_2 \in U$ with $\vec{v}_2 \notin \operatorname{Span}\{\vec{v}_1\}$. It follows from the Dependency Theorem that $\{\vec{v}_1, \vec{v}_2\}$ is linearly independent and we have that $\operatorname{Span}\{\vec{v}_1, \vec{v}_2\} \subseteq U$ since U is closed under linear combinations. If $U = \operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$, then we are done.

Otherwise, $U \neq \operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$ so there exists a $\vec{v}_3 \in U$ with $\vec{v}_3 \notin \operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$. It follows from the Dependency Theorem that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent and we have that $\operatorname{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \subseteq U$ since U is closed under linear combinations. If $U = \operatorname{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, then we are done.

Otherwise, we continue in this fashion, producing a sequence of linearly independent subsets

$$\{\vec{v}_1\}, \{\vec{v}_1, \vec{v}_2\}, \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}, \ldots$$

of U until we find a k such that $U = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$, at which point we will be done.

We are guaranteed that such a k does not exceed n since Corollary 4.3.7 shows that the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is linearly dependent whenever $k > n$.

Indeed, if $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-1}, \vec{v}_n\} \subseteq U$ is linearly independent, Theorem 4.3.4 gives that $\operatorname{rank}\left(\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n-1} & \vec{v}_n \end{bmatrix}\right) = n$, and Theorem 4.1.10 then shows that

$$\operatorname{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-1}, \vec{v}_n\} = \mathbb{R}^n.$$

Thus we have that $\mathbb{R}^n \subseteq U$. Trivially, $U \subseteq \mathbb{R}^n$, and so we have $U = \mathbb{R}^n$, that is, $U = \operatorname{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-1}, \vec{v}_n\}$. $\square$

Together, Theorem 4.4.7 and Theorem 4.4.8 show that it is *exactly* the subspaces of $\mathbb{R}^n$ that have spanning sets.

Example 4.4.9

If we examine the subspace

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; x_1 + x_2 = 0 \text{ and } x_2 - x_3 = 0 \right\}$$

of $\mathbb{R}^3$ given in Example 4.4.6, we notice that it is precisely the solution set to the system

$$\begin{array}{ccccccc} x_1 & + & x_2 & & & = & 0 \\ & & x_2 & - & x_3 & = & 0 \end{array}.$$

Solving this in the usual way, we find that the solutions are given by

$$\vec{x} = t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

Thus, $U = \operatorname{Span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

In the next section we will explore the problem of finding a spanning set for a given subspace in more detail.

Section 4.4 Problems

4.4.1. For each of the following subsets $U \subseteq \mathbb{R}^3$, determine whether or not it is a subspace of $\mathbb{R}^3$. If it is a subspace, prove it. If it is not a subspace, determine which of properties S1, S2 and S3 from Definition 4.4.1 fail.

(a) $U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \,\middle|\, x_1 + x_2 \geq 0 \right\}$.

(b) $U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \,\middle|\, x_1^2 + x_2^2 = x_3^2 \right\}$.

(c) $U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \,\middle|\, 10x_1 - 12x_2 = 3x_3 \right\}$.

(d) $U = \left\{ \vec{x} \in \mathbb{R}^3 \,\middle|\, A\vec{x} = B\vec{x} \right\}$ where $A, B \in M_{3\times 3}(\mathbb{R})$ are arbitrary matrices.

(e) $U = \left\{ a \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \in \mathbb{R}^3 \,\middle|\, a, b \in \mathbb{R} \right\}$.

4.4.2. Let U_1 and U_2 be subspaces $\mathbb{R}^n$. Define their *intersection* $U_1 \cap U_2$ and their *union* $U_1 \cup U_2$ as follows:

$$U_1 \cap U_2 = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} \in U_1 \text{ and } \vec{x} \in U_2\},$$
$$U_1 \cup U_2 = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} \in U_1 \text{ or } \vec{x} \in U_2\}.$$

For each of the following statements, either show they are true, or provide an example which shows they are false.

(a) $U_1 \cap U_2$ is a subspace of $\mathbb{R}^n$.

(b) $U_1 \cup U_2$ is a subspace of $\mathbb{R}^n$.

4.4.3. Let U be a subspace of $\mathbb{R}^n$. Define

$$U^\perp = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} \cdot \vec{s} = 0 \text{ for every } \vec{s} \in U\}.$$

Note that $U^\perp$ is read as "U perp" and is the set of vectors in $\mathbb{R}^n$ that are perpendicular to every vector belonging to U.

(a) Let $U = \operatorname{Span}\{\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]\} \subseteq \mathbb{R}^2$. Determine $U^\perp$.

(b) Show that $U^\perp$ is a subspace of $\mathbb{R}^n$.

(c) Show that $U \cap U^\perp = \{\vec{0}\}$.

4.4.4. Recall Definition 4.4.1.

(a) Give an example of a subset of $\mathbb{R}^3$ for which S1 and S2 hold, but for which S3 does not hold.

(b) Give an example of a subset of $\mathbb{R}^3$ for which S1 and S3 hold, but for which S2 does not hold.

(c) Show that there cannot be an example of a *nonempty* subset of $\mathbb{R}^3$ for which S2 and S3 hold, but for which S1 does not hold.

4.4.5. Give a geometric description of all subspaces of:

(a) $\mathbb{R}^1$.

(b) $\mathbb{R}^2$.

(c) $\mathbb{R}^3$.

Hint: Use Theorem 4.4.8.

4.5 Bases and Dimension

Having discussed the notions of spanning sets in Section 4.1 and Section 4.2, linear independence in Section 4.3 and subspaces in Section 4.4, this section will make clear the main goal of this chapter: obtaining a linearly independent spanning set for a given subspace.

4.5.1 Bases of Subspaces

At the end Section 4.4, we learned that every subspace U of $\mathbb{R}^n$ can be expressed as the span of a finite set of vectors:

$$U = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} \text{ for some } \vec{v}_1, \ldots, \vec{v}_k \in U.$$

We begin with the crucial observation that some spanning sets are "better" than others. Indeed, as we learned in Section 4.2 and Section 4.3, we can remove linear dependencies from a spanning set without affecting the resulting span. These concepts are reviewed in the following example.

Example 4.5.1 Consider the subspace

$$U = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} \right\}$$

of $\mathbb{R}^3$. Describe U geometrically.

Solution: The given spanning set for U is

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 3 \\ 0 \end{bmatrix} \right\},$$

so we let

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

and carry A to row echelon form. We have

$$\begin{bmatrix} 1 & -2 & 1 & 2 & 6 \\ 0 & 0 & 1 & 1 & 3 \\ -1 & 2 & 1 & 0 & 0 \end{bmatrix} \underset{R_3+R_1}{\longrightarrow} \begin{bmatrix} 1 & -2 & 1 & 2 & 6 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 2 & 2 & 6 \end{bmatrix} \underset{R_3-2R_2}{\longrightarrow} \begin{bmatrix} 1 & -2 & 1 & 2 & 6 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since we have leading entries in the first and third columns of a row echelon form of A, the Extraction Theorem gives that

$$S' = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is a linearly independent subset of S with $U = \operatorname{Span} S'$. It is now clear that U is a plane through the origin in $\mathbb{R}^3$.

In Example 4.5.1, both

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} \right\} \quad \text{and} \quad S' = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

are spanning sets for U, but we see that S' is a "better" spanning set since it allows us to understand what U is geometrically. This follows from the fact that S' is linearly independent, which motivates our next definition.

Definition 4.5.2
Basis

Let U be a subspace of $\mathbb{R}^n$ and let B be a finite subset[4] of U. We say that B is a **basis** for U if

(a) B is linearly independent, and

(b) $U = \operatorname{Span} B$.

Recall that we have adopted the convention that $\operatorname{Span} \varnothing = \{\vec{0}\}$ in Definition 4.1.1 and it follows from our discussion at the very end of Section 4.3 that $\varnothing$ is linearly independent. Thus $\varnothing$ is a basis for the trivial subspace $U = \{\vec{0}\}$.

We may think of a basis B for a subspace U of $\mathbb{R}^n$ as

- a *minimal spanning set for* U in the sense that B spans U, but removing even one vector from B would leave us with a set that no longer spans U. This is because B is linearly independent, so no vector in B is a linear combination of the other vectors in B by the Dependency Theorem. It then follows from the Reduction Theorem that removing a vector from B would result in a set that does not span U.

- a *maximal linearly independent subset of* U in the sense that B is linearly independent, but adding even one additional vector from U to the set B would result in a linearly dependent set. This is because B spans U, so any vector $\vec{v} \in U$ can be expressed as a linear combination of the vectors in B. Adding this vector $\vec{v}$ to the set B would create a linearly dependent set by the Dependency Theorem.

It is natural to ask if every subspace U of $\mathbb{R}^n$ has a basis, but we have in fact already answered this question:

- for the trivial subspace $U = \{\vec{0}\}$ of $\mathbb{R}^n$, it was explained right after Definition 4.5.2 that $\varnothing$ is a basis for U (in fact, since $\varnothing$ is the only linearly independent subset of $\{\vec{0}\}$, it is the *only* basis for the trivial subspace).

- for any nontrivial subspace U of $\mathbb{R}^n$, note that the spanning set constructed for U in the proof Theorem 4.4.8 was linearly independent, that is, the proof of Theorem 4.4.8 actually showed that every nontrivial subspace U of $\mathbb{R}^n$ has a basis!

[4] A set is a *finite* set if the number of elements in the set is a finite number. For example, the set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a finite set since it has 3 elements, while $\mathbb{R}^n$ not a finite set since it has infinitely many elements (we say that $\mathbb{R}^n$ is an *infinite* set).

It is important to observe that in Definition 4.5.2, we refer to "a" basis rather than "the" basis. As we will see below, every non-trivial subspace of $\mathbb{R}^n$ has infinitely many bases.

Let's begin by focusing on $U = \mathbb{R}^n$ and singling out a particularly important basis.

Example 4.5.3

Show that

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^3$.

Solution: Consider the matrix

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

Clearly, $\operatorname{rank}(A) = 3$, which is the number of rows *and* the number of columns of A. Thus B spans $\mathbb{R}^3$ by Theorem 4.1.10 and B is linearly independent by Theorem 4.3.4. Hence B is a basis for $\mathbb{R}^3$.

The basis in Example 4.5.3 is known as the *standard basis* for $\mathbb{R}^3$ which we first encountered in Example 1.2.4. We similarly have the standard basis for $\mathbb{R}^n$.

Definition 4.5.4
Standard Basis for $\mathbb{R}^n$

Let $\vec{e}_1, \ldots, \vec{e}_n \in \mathbb{R}^n$ be the columns of the $n \times n$ identity matrix I. The set $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is a basis for $\mathbb{R}^n$, called the **standard basis** for $\mathbb{R}^n$.

For instance, the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively, are

$$\{\vec{e}_1, \vec{e}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \{\vec{e}_1, \vec{e}_2, \vec{e}_3\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\},$$

both of which are illustrated in Figure 4.5.1. The standard basis for $\mathbb{R}^n$ will appear frequently in Chapter 5.

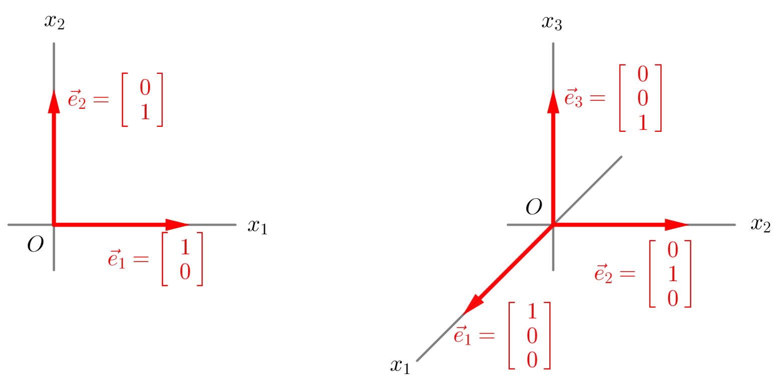


Figure 4.5.1: The standard basis for $\mathbb{R}^2$ and the standard basis for $\mathbb{R}^3$.

The next example confirms that there are bases for $\mathbb{R}^n$ other than the standard basis.

Example 4.5.5 Show that

$$B = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^2$.

Solution: Consider the matrix

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

Carrying A to row echelon form, we have

$$\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \underset{R_2 - 2R_1}{\longrightarrow} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

from which we see that $\operatorname{rank}(A) = 2$, which is both the number of rows of A and the number of columns of A. By Theorem 4.1.10, B spans $\mathbb{R}^2$ and by Theorem 4.3.4, B is linearly independent. Thus B is a basis for $\mathbb{R}^2$.

The above examples appear to indicate that a basis for $\mathbb{R}^n$ will consist of n vectors. This is indeed the case.

Theorem 4.5.6 If $B = \{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis for $\mathbb{R}^n$, then $k = n$, that is, every basis for $\mathbb{R}^n$ consists of exactly n vectors.

Proof: Consider $B = \{\vec{v}_1, \dots, \vec{v}_k\} \subseteq \mathbb{R}^n$. If B is a basis for $\mathbb{R}^n$, then B is linearly independent and $\operatorname{Span} B = \mathbb{R}^n$. Since B is linearly independent, it follows from Corollary 4.3.7 that $k \leq n$. Since $\operatorname{Span} B = \mathbb{R}^n$, it follows from Corollary 4.1.13 that $k \geq n$. Hence if B is a basis for $\mathbb{R}^n$, then $k = n$. □

Although every basis for $\mathbb{R}^n$ contains exactly n vectors, a subset of $\mathbb{R}^n$ containing exactly n vectors will not necessarily be a basis for $\mathbb{R}^n$. For example, the set

$$B = \{\vec{0}, \vec{e}_1, \vec{e}_2\} \subseteq \mathbb{R}^3$$

contains the zero vector, and is thus linearly dependent (see Example 4.3.14) and hence not a basis for $\mathbb{R}^3$ despite containing exactly 3 vectors.

Theorem 4.5.7 Let $B = \{\vec{v}_1, \dots, \vec{v}_n\} \subseteq \mathbb{R}^n$ and let $A = \begin{bmatrix} \vec{v}_1, \dots, \vec{v}_n \end{bmatrix} \in M_{n \times n}(\mathbb{R})$. Then B is a basis for $\mathbb{R}^n$ if and only if $\operatorname{rank}(A) = n$, that is, if and only if A is invertible.

Proof: Assume first that B is a basis for $\mathbb{R}^n$. Then B is linearly independent, so $\operatorname{rank}(A) = n$ by Theorem 4.3.4 since A has n columns. (We could also argue that B spans $\mathbb{R}^n$, so $\operatorname{rank}(A) = n$ by Theorem 4.1.10 since A has n rows.)

Assume now that $\operatorname{rank}(A) = n$. Then since A has n columns, B is linearly independent by Theorem 4.3.4 and since A has n rows, B spans $\mathbb{R}^n$ by Theorem 4.1.10. Thus B is a basis for $\mathbb{R}^n$.

This shows that B is a basis for $\mathbb{R}^n$ if and only if $\operatorname{rank}(A) = n$. By the Matrix Invertibility Criteria, $\operatorname{rank}(A) = n$ if and only if A is invertible. □

Example 4.5.8 Determine which of the following sets form a basis for $\mathbb{R}^3$.

(a) $B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$.

(b) $B_2 = \left\{ \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} \right\}$.

(c) $B_3 = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix} \right\}$.

(d) $B_4 = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix} \right\}$.

Solution:

(a) Since B_1 contains $2 < 3$ vectors, B_1 is not a basis for $\mathbb{R}^3$ by Theorem 4.5.6.

(b) Since

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & -1 \\ 3 & 4 & 5 \end{bmatrix} \underset{\substack{R_2 - 3R_1 \\ R_3 - 3R_1}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -5 & -10 \\ 0 & -2 & -4 \end{bmatrix} \underset{R_3 - \frac{2}{5}R_2}{\longrightarrow} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -5 & -10 \\ 0 & 0 & 0 \end{bmatrix},$$

we see that $\operatorname{rank}(A) = 2 < 3$. Thus B_2 is not a basis for $\mathbb{R}^3$ by Theorem 4.5.7.

(c) Since

$$A = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 1 & 2 \\ 1 & -1 & 3 \end{bmatrix} \underset{\substack{R_2 - 2R_1 \\ R_3 - R_1}}{\longrightarrow} \begin{bmatrix} 1 & 3 & -2 \\ 0 & -5 & 6 \\ 0 & -4 & 5 \end{bmatrix} \underset{R_3 - \frac{4}{5}R_2}{\longrightarrow} \begin{bmatrix} 1 & 3 & -2 \\ 0 & -5 & 6 \\ 0 & 0 & 1/5 \end{bmatrix},$$

we see that $\operatorname{rank}(A) = 3$. Thus B_3 is a basis for $\mathbb{R}^3$ by Theorem 4.5.7.

(d) Since B_4 contains $4 > 3$ vectors, B_4 is not a basis for $\mathbb{R}^3$ by Theorem 4.5.6.

Exercise 68 Which of the following are bases for $\mathbb{R}^3$?

(a) $B_1 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$.

(b) $B_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 7 \end{bmatrix} \right\}$.

For a set $B = \{\vec{v}_1, \ldots, \vec{v}_n\} \subseteq \mathbb{R}^n$, carefully reviewing the previous examples will lead us to conjecture that B spans $\mathbb{R}^n$ exactly when B is linearly independent. The following corollary confirms observation.

Corollary 4.5.9 Let $B = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a set of n vectors in $\mathbb{R}^n$. Then B spans $\mathbb{R}^n$ if and only if B is linearly independent.

Given a set $B = \{\vec{v}_1, \ldots, \vec{v}_k\}$ of k vectors in $\mathbb{R}^n$, it is important to note that we *cannot* apply Corollary 4.5.9 when $k \neq n$. Indeed, with $n = 3$,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a linearly independent set of $k = 2 \neq n$ vectors in $\mathbb{R}^3$ that does not span $\mathbb{R}^3$ and

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is a set of $k = 4 \neq n$ vectors spanning $\mathbb{R}^3$ that is linearly dependent.

Exercise 69 Prove Corollary 4.5.9.

We now turn our attention to finding bases for subspaces U of $\mathbb{R}^n$ where $U \neq \mathbb{R}^n$.

Example 4.5.10 Find a basis for the subspace

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; x_1 - x_2 + 2x_3 = 0 \right\}$$

of $\mathbb{R}^3$.

Solution: Let $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right] \in U$. Then $x_1 - x_2 + 2x_3 = 0$, so $x_1 = x_2 - 2x_3$. We have

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

Letting

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\},$$

we see that $U \subseteq \operatorname{Span} B$. Since the vectors in B belong to U and U is closed under linear combinations by virtue being a subspace, $\operatorname{Span} B \subseteq U$. Thus $U = \operatorname{Span} B$. Since neither vector in B is a scalar multiple of the other, B is linearly independent and thus a basis for U.

In Example 4.5.10, we were not given a spanning set for U in advance, rather, we had to derive one. When determining a spanning set for a subspace U of $\mathbb{R}^n$, we choose an arbitrary $\vec{x} \in U$ and try to "decompose" $\vec{x}$ as a linear combination of some vectors $\vec{v}_1, \dots, \vec{v}_k \in U$. This shows that $U \subseteq \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\}$. Technically, we should also show that $\operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\} \subseteq U$, but this is trivial as U is a subspace and thus contains all linear combinations of $\vec{v}_1, \dots, \vec{v}_k$ (see the comments immediately following Definition 4.4.1). Thus for a subspace U of $\mathbb{R}^n$ with $\vec{v}_1, \dots, \vec{v}_k \in U$,

$$U \subseteq \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\} \Longrightarrow U = \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\},$$

and we don't normally show (or even mention) that $\operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\} \subseteq U$.

Example 4.5.11 Consider the subspace

$$U = \left\{ \begin{bmatrix} a-b \\ b-c \\ c-a \end{bmatrix} \,\middle|\, a, b, c \in \mathbb{R} \right\}$$

of $\mathbb{R}^3$. Find a basis for U.

Solution: Let $\vec{x} \in U$. Then for some $a, b, c \in \mathbb{R}$,

$$\vec{x} = \begin{bmatrix} a-b \\ b-c \\ c-a \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}.$$

Thus

$$U = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

Now since

$$\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = -\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

we have from the Reduction Theorem that

$$U = \operatorname{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

so

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a spanning set for U. Moreover, since neither vector in B is a scalar multiple of the other, B is linearly independent and hence a basis for U.

Example 4.5.12 Find a basis for the subspace

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \,\middle|\, x_1 + x_2 = 0 \text{ and } x_2 - x_3 = 0 \right\}$$

of $\mathbb{R}^3$.

Solution: Let $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right] \in U$. Then $x_1 + x_2 = 0$ and $x_2 - x_3 = 0$ and thus $x_1 = -x_2$ and $x_3 = x_2$. It follows that

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus $U = \operatorname{Span}\left\{ \left[\begin{smallmatrix} -1 \\ 1 \\ 1 \end{smallmatrix}\right] \right\}$. Hence the set

$$B = \left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is a spanning set for U. Since B consists of a single nonzero vector, B is linearly independent and is hence a basis for U.

The following algorithm summarizes the method we have used to construct a basis for a subspace U of $\mathbb{R}^n$.

ALGORITHM (Finding a basis for a subspace U of $\mathbb{R}^n$)

To find a basis for a subspace U of $\mathbb{R}^n$, perform the following steps.

- **Step 1:** Pick an arbitrary $\vec{x} \in U$ and then use the definition of U to express $\vec{x}$ as a linear combination of some vectors $\vec{v}_1, \ldots, \vec{v}_k \in U$. This gives a spanning set $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ for U.
- **Step 2:** Remove any dependencies from S by using the Extraction Theorem (or the Dependency Theorem and the Reduction Theorem) to obtain a linearly independent set $B \subseteq S$ with $\operatorname{Span} B = U$.

The resulting set B is a basis for U.

Exercise 70 Find a basis for the subspace

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \middle| \, x_1 + 2x_2 = 0 \right\}$$

of $\mathbb{R}^3$.

Example 4.5.1 justified why it is to our advantage to remove any dependencies from a spanning set of a subspace to obtain a basis. The next theorem shows another advantage of obtaining a basis.

Theorem 4.5.13 If $B = \{\vec{v}_1, \ldots, \vec{v}_k\}$ is a basis for a subspace $U \subseteq \mathbb{R}^n$, then every $\vec{x} \in U$ can be expressed as a linear combination $\vec{v}_1, \ldots, \vec{v}_k$ in a unique way.

Proof: Since B is a basis for U, B is linearly independent and B spans U. Since $U = \operatorname{Span} B$, every $\vec{x} \in U$ can be expressed as a linear combination of the vectors in B. Thus we are left to show that this expression is unique. Let $\vec{x} \in U$ and suppose we have two such expressions

$$\vec{x} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \quad \text{and} \quad \vec{x} = d_1\vec{v}_1 + \cdots + d_k\vec{v}_k$$

for some $c_1, d_1, \ldots, c_k, d_k \in \mathbb{R}$. We must show that $c_i = d_i$ for all $i = 1, \ldots, k$. Since both expressions are equal to $\vec{x}$, we have

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = d_1\vec{v}_1 + \cdots + d_k\vec{v}_k.$$

Rearranging gives

$$(c_1 - d_1)\vec{v}_1 + \cdots + (c_k - d_k)\vec{v}_k = \vec{0}.$$

Since B is linearly independent, we have that $c_1 - d_1 = \cdots = c_k - d_k = 0$, that is, $c_i = d_i$ for $i = 1, \ldots, k$, as desired. □

4.5.2 Dimension of a Subspace

We now formally define the dimension of a subspace. Intuitively we have an understanding of what dimension is: a line is 1-dimensional and a plane is 2-dimensional. Through the many examples we have seen in this chapter, we have repeatedly noticed that a line can be spanned by a single vector and a plane can be spanned by two vectors. With the notion of basis, we now understand that a line through the origin has a basis containing one vector and a plane through the origin has a basis containing two vectors. This section will use bases to formally define the dimension of a subspace. This notion of dimension can be extended to sets that are not subspaces of $\mathbb{R}^n$, but we will not pursue that idea here.

Motivated by our observations with lines and planes, we want to define the dimension of a subspace U to be the number of vectors in any basis of U. For this to be logically sound, we need to be sure that any two bases of U contain the same number of vectors. This will follow from our next two theorems.

Theorem 4.5.14 Let $B = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a basis for a subspace U of $\mathbb{R}^n$. If $C = \{\vec{w}_1, \dots, \vec{w}_\ell\}$ is a set in U with $\ell > k$, then C is linearly dependent.

Proof: We prove Theorem 4.5.14 in the case $k = 2$ and $\ell = 3$, the proof of the general result being similar. Thus we assume $B = \{\vec{v}_1, \vec{v}_2\}$ is a basis for U and that $C = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is a set of three vectors in U. Since B is a basis for U, Theorem 4.5.13 gives that there are unique $a_1, a_2, b_1, b_2, c_1, c_2 \in \mathbb{R}$ so that

$$\vec{w}_1 = a_1\vec{v}_1 + a_2\vec{v}_2, \quad \vec{w}_2 = b_1\vec{v}_1 + b_2\vec{v}_2 \quad \text{and} \quad \vec{w}_3 = c_1\vec{v}_1 + c_2\vec{v}_2.$$

Now for $t_1, t_2, t_3 \in \mathbb{R}$, consider

$$\begin{aligned}\vec{0} &= t_1\vec{w}_1 + t_2\vec{w}_2 + t_3\vec{w}_3\\ &= t_1(a_1\vec{v}_1 + a_2\vec{v}_2) + t_2(b_1\vec{v}_1 + b_2\vec{v}_2) + t_3(c_1\vec{v}_1 + c_2\vec{v}_2)\\ &= (a_1t_1 + b_1t_2 + c_1t_3)\vec{v}_1 + (a_2t_1 + b_2t_2 + c_2t_3)\vec{v}_2\end{aligned}$$

Since $B = \{\vec{v}_1, \vec{v}_2\}$ is linearly independent, we have

$$\begin{array}{ccccccc} a_1t_1 & + & b_1t_2 & + & c_1t_3 & = & 0 \\ a_2t_1 & + & b_2t_2 & + & c_2t_3 & = & 0 \end{array}.$$

This is an underdetermined homogeneous system, so it is consistent with nontrivial solutions and it follows that $C = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is linearly dependent. □

It follows from Theorem 4.5.14 that if $B = \{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis for a subspace U of $\mathbb{R}^n$ and $C = \{\vec{w}_1, \dots, \vec{w}_\ell\}$ is a linearly *independent* subset of U, then $\ell \leq k$. We now state our main result.

Theorem 4.5.15 If $B = \{\vec{v}_1, \dots, \vec{v}_k\}$ and $C = \{\vec{w}_1, \dots, \vec{w}_\ell\}$ are bases for a subspace U of $\mathbb{R}^n$, then $k = \ell$.

Proof: Since B is a basis for U and C is linearly independent, we have that $\ell \leq k$. Since C is a basis for U and B is linearly independent, $k \leq \ell$. Hence $k = \ell$. □

Hence, given a subspace U of $\mathbb{R}^n$, there may be many bases for U, but they will all contain the same number of vectors. This allows us to make the following definition.

Definition 4.5.16
Dimension

If $B = \{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis for a subspace U of $\mathbb{R}^n$, then we say the **dimension** of U is k, and we write $\dim(U) = k$.

If $U = \{\vec{0}\}$, then $\dim(U) = 0$ since $\varnothing$ is a basis for U.

Example 4.5.17 Since the standard basis for $\mathbb{R}^n$ is $\{\vec{e}_1, \dots, \vec{e}_n\}$, we see that $\dim(\mathbb{R}^n) = n$.

Example 4.5.18 We saw in Example 4.5.11 that the subspace

$$U = \left\{ \begin{bmatrix} a-b \\ b-c \\ c-a \end{bmatrix} \,\middle|\, a, b, c \in \mathbb{R} \right\}$$

of $\mathbb{R}^3$ had basis

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\},$$

so $\dim(U) = 2$.

The following theorem shows why it can be useful to know the dimension of a subspace of $\mathbb{R}^n$. Note that this theorem generalizes Corollary 4.3.7, Corollary 4.1.13 and Corollary 4.5.9 (respectively) to arbitrary subspaces of $\mathbb{R}^n$.

Theorem 4.5.19 If U is a k-dimensional subspace of $\mathbb{R}^n$ with $k > 0$, then

(a) A set of more than k vectors in U is linearly dependent,

(b) A set of fewer than k vectors in U cannot span U,

(c) A set of exactly k vectors in U spans U if and only if it is linearly independent.

Proof: Let U be a k-dimensional subspace of $\mathbb{R}^n$ with $k > 0$. Then any basis for U contains exactly k vectors.

(a) If a subset C of U has more than k vectors, then Theorem 4.5.14 shows that C is linearly dependent.

(b) Let C be a subset of U with fewer than k vectors and suppose that C spans U. If necessary, we remove any dependencies from C using the Dependency Theorem and the Reduction Theorem to obtain a basis for U that contains less than k vectors which implies that $\dim(U) < k$. This contradicts $\dim(U) = k$. Thus, if C has fewer than k vectors, then C cannot span U.

(c) Let B be a subset of U with k vectors. Assume first that B spans U. We must show that B is linearly independent. Suppose instead that B is linearly dependent. Then by the Dependency Theorem and the Reduction Theorem, there is a subset C of B with less than k vectors that also spans U which contradicts (b). Thus B must be linearly independent.

    Now assume that B is linearly independent. We must show that B spans U. Suppose instead that B does not span U. Then there is an $\vec{x} \in U$ with $\vec{x} \notin \operatorname{Span} B$. By the Dependency Theorem, the set $C = B \cup \{\vec{x}\}$ is a linearly independent subset of U with $k + 1$ vectors which contradicts (a). Thus B must span U. $\square$

If we know the dimension of a subspace U of $\mathbb{R}^n$, then Theorem 4.5.19(c) makes it easier to determine if a set $B = \{\vec{v}_1, \ldots, \vec{v}_k\}$ of k vectors belonging to U forms a basis for U, since in this case, we need only verify that B spans U *or* that B is linearly independent.

Example 4.5.20

Let U be a subspace of $\mathbb{R}^3$ with $\dim(U) = 2$. Let $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ be vectors in U. Show that $\{\vec{v}_1, \vec{v}_2\}$ is a basis for U.

Solution: Since neither $\vec{v}_1$ nor $\vec{v}_2$ are scalar multiples of the other, we have that $\{\vec{v}_1, \vec{v}_2\}$ is a linearly independent set of two vectors in U. Since $\dim(U) = 2$, we have that $U = \operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$ by Theorem 4.5.19(c). Thus $\{\vec{v}_1, \vec{v}_2\}$ is a basis for U.

Note that we must know $\dim(U)$ *before* we use Theorem 4.5.19. In the previous example, we could not have used the linear independence of $\{\vec{v}_1, \vec{v}_2\}$ to conclude that $U = \operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$ if we weren't given the dimension of U.

4.5.3 Extending a Basis

Now that we are able to find a basis for for subspaces of $\mathbb{R}^n$, we would like to *extend* this basis to a obtain a basis for $\mathbb{R}^n$. More specifically, Given a k-dimensional subspace U of $\mathbb{R}^n$ with basis $B = \{\vec{v}_1, \ldots, \vec{v}_k\}$, we would like to find $n - k$ vectors $\vec{u}_{k+1}, \ldots, \vec{u}_n$ so that

$$B' = \{\vec{v}_1, \ldots, \vec{v}_k, \vec{u}_{k+1}, \ldots, \vec{u}_n\}$$

is a basis for $\mathbb{R}^n$.

Example 4.5.21

Let

$$B = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \\ -1 \end{bmatrix} \right\}$$

be a basis for a subspace U of $\mathbb{R}^4$. Extend B to a basis B' of $\mathbb{R}^4$.

Solution: We first construct a spanning set S for $\mathbb{R}^4$ that contains B. Consider

$$S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Then S is clearly a spanning set for $\mathbb{R}^4$ since the last four vectors in S are the standard basis vectors for $\mathbb{R}^4$. Since

$$\begin{bmatrix} 1 & 3 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & -2 \\ 0 & 0 & 1 & 2 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

we see that there are leading entries in the first, second, third and fifth columns of the reduced row echelon form. Thus the Extraction Theorem gives that

$$B' = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is an extension of B to a basis for $\mathbb{R}^4$.

Example 4.5.21 shows that we use the same method that we developed to extract a basis from a spanning set. Given a basis B for a k-dimensional subspace U of $\mathbb{R}^n$, we construct a matrix $A \in M_{n\times(k+n)}(\mathbb{R})$ with the k basis vectors of B as the first k columns and the n vectors from any basis for $\mathbb{R}^n$ as the last n columns. We then carry A to any row echelon form R (note that we carried A to reduced row echelon form in Example 4.5.21). Since the first k columns of A are the basis vectors for U, they are linearly independent, and so the first k columns of R (which correspond to the first k columns in A, namely the vectors in B) will have leading entries in them. The remaining n columns of R will contain an additional $n-k$ leading entries. These last $n-k$ columns with leading entries will correspond to those columns in A that we must add to B to create B'.

The order in which we add columns to A is important. Given that we are extending a basis B for a k-dimensional subspace U of $\mathbb{R}^n$ to a basis B' for $\mathbb{R}^n$, we must take the k vectors in B as the *first* k columns of A. Indeed, if we had chosen A to be the matrix

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

in Example 4.5.21, then we would have seen that A is already in reduced row echelon form. Our algorithm would say to take the first four columns of A as our basis B', which in this case would be the standard basis. Although this is a basis for $\mathbb{R}^4$, it doesn't contain any of the vectors from B, and is thus not an extension of B to a basis for $\mathbb{R}^4$.

ALGORITHM (Extending a Basis for a Subspace of $\mathbb{R}^n$ to a Basis for $\mathbb{R}^n$)

Given a subspace U of $\mathbb{R}^n$ with basis $B = \{\vec{v}_1, \ldots, \vec{v}_k\}$, we can extend B to a basis B' of $\mathbb{R}^n$ by performing the following steps.

- **Step 1:** Construct a matrix $A \in M_{n\times(k+n)}(\mathbb{R})$ whose first k columns are the vectors in B and whose last n columns are the standard basis vectors (or any n vectors that form a basis for $\mathbb{R}^n$).

- **Step 2:** Carry A to any row echelon form and use the Extraction Theorem to obtain a linearly independent set B' with $B \subseteq B'$ and $\operatorname{Span} B' = \mathbb{R}^n$.

The resulting set B' is an extension of B to a basis for $\mathbb{R}^n$.

Exercise 71 Let $U = \operatorname{Span} S$, where

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \\ -1 \end{bmatrix} \right\}.$$

(a) Find a basis B for U with $B \subseteq S$.

(b) Extend B to a basis B' for $\mathbb{R}^4$.

Section 4.5 Problems

4.5.1. Determine which of the following are bases for $\mathbb{R}^2$.

(a) $B_1 = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \end{bmatrix} \right\}$. (b) $B_2 = \left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -9 \end{bmatrix} \right\}$. (c) $B_3 = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \end{bmatrix} \right\}$.

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

(a) $B_1 = \left\{ \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right\}$. (b) $B_2 = \left\{ \begin{bmatrix} 1 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \right\}$.

(c) $B_3 = \left\{ \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}$.

4.5.3. Let $\vec{y}, \vec{z} \in \mathbb{R}^n$ and consider the set $U = \{ \vec{x} \in \mathbb{R}^n \mid \vec{x} \cdot \vec{y} = 0 \text{ and } \vec{x} \cdot \vec{z} = 0 \}$.

(a) Show that U is a subspace of $\mathbb{R}^n$.

(b) Find a basis for U given that $n = 3$ and

$$\vec{y} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix}.$$

What is $\dim(U)$?

4.5.4. Let $\vec{y}, \vec{z} \in \mathbb{R}^n$ and consider the set $U = \{ \vec{x} \in \mathbb{R}^n \mid \vec{x} \cdot \vec{y} = \vec{x} \cdot \vec{z} \}$.

(a) Show that U is a subspace of $\mathbb{R}^n$.

(b) Find a basis for U given that $n = 3$ and

$$\vec{y} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix}.$$

What is $\dim(U)$?

4.5.5. Let $B = \{\vec{v}_1, \vec{v}_2\}$ be a basis for a subspace U of $\mathbb{R}^4$ and let $\vec{w}_1, \vec{w}_2, \vec{w}_3 \in U$.

(a) Prove that $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is linearly dependent. (Give two arguments: one involving dimension, and one not relying on dimension.)

(b) Find three distinct vectors $\vec{x}_1, \vec{x}_2, \vec{x}_3 \in U$ so that $U = \operatorname{Span}\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$.

(c) Find three distinct vectors $\vec{y}_1, \vec{y}_2, \vec{y}_3 \in U$ so that $U \neq \operatorname{Span}\{\vec{y}_1, \vec{y}_2, \vec{y}_3\}$.

4.5.6. Consider the set $C = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\}$, where

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 4 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 6 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 2 \\ 1 \\ -6 \\ 6 \end{bmatrix}, \quad \vec{v}_4 = \begin{bmatrix} -1 \\ 2 \\ 5 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{v}_5 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

and define $U = \operatorname{Span} C$.

(a) Find a basis B for U with $B \subseteq C$. What is the dimension of U?

(b) Extend your basis B from part (a) to a basis B' for $\mathbb{R}^4$.

4.6 Fundamental Subspaces Associated with a Matrix

Having completed our study of subspaces and their bases, we now examine two important subspaces that are related to a matrix $A \in M_{m\times n}(\mathbb{R})$. We begin with a couple definitions.

Definition 4.6.1
Nullspace

Let $A \in M_{m\times n}(\mathbb{R})$. The **nullspace** of A (sometimes called the **kernel** of A) is the subset of $\mathbb{R}^n$ defined by

$$\text{Null}(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}.$$

Note that Null(A) is simply the set of solutions to the homogeneous system of linear equations $A\vec{x} = \vec{0}$.

Definition 4.6.2
Column Space

Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{m\times n}(\mathbb{R})$. The **column space** of A is the subset of $\mathbb{R}^m$ defined by

$$\text{Col}(A) = \{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\} = \text{Span}\{\vec{a}_1, \ldots, \vec{a}_n\}.$$

Simply put, Col(A) is the set of all linear combinations of the columns of A. The equality

$$\{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\} = \text{Span}\{\vec{a}_1, \ldots, \vec{a}_n\}$$

may appear odd at first glance, but recall the matrix–vector product: for $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, we have

$$A\vec{x} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n$$

which is a linear combination of the columns of A. Thus, if we compute $A\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$, then we computed have all linear combinations of the columns of A which gives us Span$\{\vec{a}_1, \ldots, \vec{a}_n\}$.

Theorem 4.6.3

Let $A \in M_{m\times n}(\mathbb{R})$. Then Null$(A)$ is a subspace of $\mathbb{R}^n$ and Col(A) is a subspace of $\mathbb{R}^m$.

Proof: We first show Null(A) is a subspace of $\mathbb{R}^n$. Since $A\vec{0}_{\mathbb{R}^n} = \vec{0}_{\mathbb{R}^m}$, $\vec{0}_{\mathbb{R}^n} \in \text{Null}(A)$. For $\vec{y}, \vec{z} \in \text{Null}(A)$, we have that $A\vec{y} = \vec{0}_{\mathbb{R}^m} = A\vec{z}$. Then

$$A(\vec{y} + \vec{z}) = A\vec{y} + A\vec{z} = \vec{0}_{R^m} + \vec{0}_{\mathbb{R}^m} = \vec{0}_{R^m}$$

so $\vec{y} + \vec{z} \in \text{Null}(A)$. For $c \in \mathbb{R}$,

$$A(c\vec{x}) = cA\vec{x} = c(\vec{0}_{\mathbb{R}^m}) = \vec{0}_{\mathbb{R}^m}$$

so $c\vec{x} \in \text{Null}(A)$. Thus Null$(A)$ is a subspace of $\mathbb{R}^n$.

Letting $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in M_{m\times n}(\mathbb{R})$, we have that Col$(A) = \text{Span}\{\vec{a}_1, \ldots, \vec{a}_n\}$ is a subspace of $\mathbb{R}^m$ by Theorem 4.4.7. $\square$

Having shown that Null(A) and Col(A) are subspaces, it is natural to seek bases for these subspaces. We begin with the nullspace.

Example 4.6.4 Let

$$A = \begin{bmatrix} 1 & 1 & 3 & 4 \\ 1 & -1 & -1 & 0 \\ -1 & 3 & 5 & 4 \\ 2 & 1 & 4 & 6 \end{bmatrix}.$$

Find a basis for $\mathrm{Null}(A)$.

Solution: Since $\mathrm{Null}(A)$ is the set of solutions to the homogeneous system of linear equations $A\vec{x} = \vec{0}$, we begin by solving the system:

$$\begin{bmatrix} 1 & 1 & 3 & 4 \\ 1 & -1 & -1 & 0 \\ -1 & 3 & 5 & 4 \\ 2 & 1 & 4 & 6 \end{bmatrix} \underset{\substack{R_2-R_1 \\ R_3+R_1 \\ R_4-2R_1}}{\longrightarrow} \begin{bmatrix} 1 & 1 & 3 & 4 \\ 0 & -2 & -4 & -4 \\ 0 & 4 & 8 & 8 \\ 0 & -1 & -2 & -2 \end{bmatrix} \underset{-\frac{1}{2}R_2}{\longrightarrow} \begin{bmatrix} 1 & 1 & 3 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 4 & 8 & 8 \\ 0 & -1 & -2 & -2 \end{bmatrix} \underset{\substack{R_3-4R_2 \\ R_4+R_2}}{\overset{R_1-R_2}{\longrightarrow}} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus we see that

$$\vec{x} = \begin{bmatrix} -s-2t \\ -2s-2t \\ s \\ t \end{bmatrix} = s\begin{bmatrix} -1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -2 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \quad s, t, \in \mathbb{R}.$$

Letting

$$B = \left\{ \begin{bmatrix} -1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\},$$

we see that $\mathrm{Null}(A) \subseteq \mathrm{Span}\, B$ (and that $\mathrm{Span}\, B \subseteq \mathrm{Null}(A)$ since $B \subseteq \mathrm{Null}(A)$ and $\mathrm{Null}(A)$ is closed under linear combinations). Thus $\mathrm{Null}(A) = \mathrm{Span}\, B$ and so B is a spanning set for $\mathrm{Null}(A)$. Since each vector in B has a 1 where the other has a 0, the set B is linearly independent, and hence a basis for $\mathrm{Null}(A)$.

We make a couple of observations regarding our solution to Example 4.6.4. First, notice that by carrying A to (reduced) row echelon form, we obtain the vector equation of the solution to the homogeneous system of linear equations. This immediately gives us a spanning set for $\mathrm{Null}(A)$.

Secondly, since our spanning set B has just two vectors, it was likely expected that our justification for linear independence would have been something akin to "since neither of the vectors in B is a scalar multiple of the other, B is linearly independent" which would be a correct justification. Instead, however, we chose to argue the linear independence of B by saying that each vector in B has a 1 where the other has a 0. The reason for this is that the latter argument will extend to cases when our spanning set for $\mathrm{Null}(A)$ contains more than two vectors. For example, consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix},$$

which is already in reduced row echelon form. Solving the homogeneous system of linear equations $A\vec{x} = \vec{0}$ shows that x_2, x_3 and x_5 are free variables so the solution is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -t_1 - t_2 - 4t_3 \\ 1t_1 \\ 1t_2 \\ -2t_3 \\ 1t_4 \end{bmatrix} = \begin{bmatrix} -t_1 - t_2 - 4t_3 \\ 1t_1 + 0t_2 + 0t_3 \\ 0t_1 + 1t_2 + 0t_3 \\ 0t_1 + 0t_2 - 2t_3 \\ 0t_1 + 0t_2 + 1t_4 \end{bmatrix} = t_1 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t_3 \begin{bmatrix} -4 \\ 0 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

for $t_1, t_2, t_3 \in \mathbb{R}$. We see that

$$B = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right\}$$

is a spanning set for Null(A). Since each vector has a 1 where the others have a 0, no vector in B is a linear combination of the others, so the Dependency Theorem gives that B is linearly independent and thus B is a basis for Null(A). We see that the spanning set for Null(A) generated by solving $A\vec{x} = \vec{0}$ via reducing A to reduced row echelon form will always be linearly independent! Thus, once we solve $A\vec{x} = \vec{0}$, we can simply write down the basis for Null(A) without any further comment.

It is also worth reminding the reader that if $A\vec{x} = \vec{0}$ has only the trivial solution, then Null$(A) = \{\vec{0}\}$ so $\varnothing$ is a basis for Null(A).

Exercise 72 Let

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

Find a basis for Null(A).

We now turn our attention to finding a basis for Col(A).

Example 4.6.5 Let

$$A = \begin{bmatrix} 1 & 1 & 3 & 4 \\ 1 & -1 & -1 & 0 \\ -1 & 3 & 5 & 4 \\ 2 & 1 & 4 & 6 \end{bmatrix}.$$

Find a basis for Col(A).

Solution: Let

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 5 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 4 \\ 6 \end{bmatrix} \right\}.$$

Then by definition, Col$(A) = \operatorname{Span} S$. Thus we only need to check S for linear independence and remove any dependencies in order to obtain a basis for Col(A). We know from Example 4.6.4 that

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

is the reduced row echelon form of A. Since there are leading entries in the first and second column of R, the Extraction Theorem gives that

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 3 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathrm{Col}(A)$.

Example 4.6.5 should feel familiar - we have seen it before in Example 4.3.11, but there we were not considering the vectors in S as the columns of a matrix. Given a subspace $U = \mathrm{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$ of $\mathbb{R}^n$, we can define a matrix $A = \begin{bmatrix} \vec{v}_1, \ldots, \vec{v}_k \end{bmatrix} \in M_{n\times k}(\mathbb{R})$, so that we have $U = \mathrm{Col}(A)$. It follows that *every* subspace U of $\mathbb{R}^n$ is the column space of some matrix.

Exercise 73 Let

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 \\ 3 & 5 & 2 & 4 \end{bmatrix}$$

Find a basis for $\mathrm{Col}(A)$.

It is clear that the nullspace of a matrix is closely tied to homogeneous systems of linear equations. The next theorem shows how the columns space of a matrix is also related to systems of linear equations.

Theorem 4.6.6 Let $A \in M_{m\times n}(\mathbb{R})$ and let $\vec{b} \in \mathbb{R}^m$. Then the system of linear equations $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b} \in \mathrm{Col}(A)$.

Proof: By part (a) of Theorem 3.3.9, $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b}$ can be expressed as a linear combination of the columns of A, which is equivalent to $\vec{b} \in \mathrm{Col}(A)$ by Definition 4.6.2. □

Note that by using the Extraction Theorem, our method to find a basis for $\mathrm{Col}(A)$ only requires A to be carried to row echelon form. However, since we will often find bases for $\mathrm{Null}(A)$ and $\mathrm{Col}(A)$ together, it is natural to carry A to reduced row echelon form.

Example 4.6.7 Let

$$A = \begin{bmatrix} 1 & 2 & 1 & 3 & 4 \\ 3 & 6 & 2 & 6 & 9 \\ -2 & -4 & 1 & 1 & -1 \end{bmatrix}.$$

Find a basis for $\mathrm{Null}(A)$ and $\mathrm{Col}(A)$, and state the dimensions of these subspaces.

Solution: Carrying A to reduced row echelon form gives

$$\begin{bmatrix} 1 & 2 & 1 & 3 & 4 \\ 3 & 6 & 2 & 6 & 9 \\ -2 & -4 & 1 & 1 & -1 \end{bmatrix} \underset{\substack{R_2-3R_1 \\ R_3+2R_1}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 1 & 3 & 4 \\ 0 & 0 & -1 & -3 & -3 \\ 0 & 0 & 3 & 7 & 7 \end{bmatrix} \overset{R_1+R_2}{\underset{R_3+3R_2}{\longrightarrow}}$$
$$\begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & -1 & -3 & -3 \\ 0 & 0 & 0 & -2 & -2 \end{bmatrix} \underset{\substack{-R_2 \\ -\frac{1}{2}R_3}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \underset{R_2-3R_3}{\longrightarrow} \begin{bmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Solving $A\vec{x} = \vec{0}$, we have

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}$$

so

$$B_1 = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathrm{Null}(A)$ showing that $\dim(\mathrm{Null}(A)) = 2$. As the first, third and fourth columns of a row echelon form of A have leading entries, the Extraction Theorem gives that

$$B_2 = \left\{ \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathrm{Col}(A)$ and $\dim(\mathrm{Col}(A)) = 3$.

Given a matrix $A \in M_{m\times n}(\mathbb{R})$, the number of vectors in a basis for $\mathrm{Null}(A)$ will be the number of free variables in the solution to the homogeneous system of linear equations $A\vec{x} = \vec{0}$. By part (b) of the System–Rank Theorem, there are $n - \mathrm{rank}(A)$ parameters. Thus $\dim(\mathrm{Null}(A)) = n - \mathrm{rank}(A)$. We make the following definition.

Definition 4.6.8
Nullity

Let $A \in M_{m\times n}(\mathbb{R})$. The **nullity** of A, denoted by $\mathrm{nullity}(A)$ is defined by

$$\mathrm{nullity}(A) = n - \mathrm{rank}(A).$$

It follows from Definition 4.6.8 that $\dim(\mathrm{Null}(A)) = \mathrm{nullity}(A)$. The number of vectors in a basis for $\mathrm{Col}(A)$ will be the number of columns with leading entries in any row echelon form of A, that is, $\dim(\mathrm{Col}(A)) = \mathrm{rank}(A)$. This verifies the following theorem.

Theorem 4.6.9 Let $A \in M_{m\times n}(\mathbb{R})$. Then

(a) $\dim(\mathrm{Null}(A)) = \mathrm{nullity}(A)$.

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

We also have the following result, known as the Rank-Nullity Theorem.

Theorem 4.6.10 **(Rank–Nullity Theorem)**

For any $A \in M_{m\times n}(\mathbb{R})$,

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

Proof: Let $A \in M_{m\times n}(\mathbb{R})$. Using Definition 4.6.8, we have

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = \mathrm{rank}(A) + n - \mathrm{rank}(A) = n. \qquad \square$$

It follows from the Rank-Nullity Theorem that for any $A \in M_{m\times n}(\mathbb{R})$,

$$\dim(\mathrm{Null}(A)) + \dim(\mathrm{Col}(A)) = n.$$

This will have a meaningful interpretation in Chapter 5.

Section 4.6 Problems

4.6.1. For each matrix A given below, find bases for $\text{Null}(A)$ and $\text{Col}(A)$, and state the dimensions of each of these subspaces. Also show that each matrix satisfies the Rank–Nullity Theorem.

(a) $A = \begin{bmatrix} 2 & -4 & 5 \\ 1 & -2 & 2 \\ -3 & 6 & -7 \end{bmatrix}$.

(b) $A = \begin{bmatrix} 1 & 1 & 5 & 1 \\ 1 & 2 & 7 & 2 \\ 2 & 3 & 12 & 3 \end{bmatrix}$.

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

(d) A is any invertible 4×4 matrix.

4.6.2. Let
$$A = \begin{bmatrix} 1 & -2 & 1 & 2 \\ 2 & -4 & 3 & 5 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

(a) Determine a basis B_1 for $\text{Null}(A)$ and a basis C_1 for $\text{Col}(A)$.

(b) Determine a basis C_2 for $\text{Null}(A^T)$ and a basis B_2 for $\text{Col}(A^T)$.

(c) Show that $B = B_1 \cup B_2$ is a basis for $\mathbb{R}^4$.

(d) Show that $C = C_1 \cup C_2$ is a basis for $\mathbb{R}^3$.

4.6.3. Let $A \in M_{m\times n}(\mathbb{R})$ and $B \in M_{n\times k}(\mathbb{R})$.

(a) Show that $\text{Null}(B) \subseteq \text{Null}(AB)$.

(b) Show that $\text{Col}(AB) \subseteq \text{Col}(A)$.

4.6.4. Let $A \in M_{n\times n}(\mathbb{R})$. Show that if $A^2 = 0_{n\times n}$, then $\text{Col}(A) \subseteq \text{Null}(A)$.

4.7 Summary

We conclude this chapter by summarizing what we have accomplished thus far.

In Chapter 1, we learned the basic operations involving vectors: addition and scalar multiplication, which led to linear combinations. We also used vectors to construct equations for lines and planes. In Chapter 2, we learned how to solve systems of linear equations and understood that we were ultimately intersecting lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$. We also encountered the System–Rank Theorem which allowed us to use the rank of a matrix to comment on the consistency and number of solutions of a system of linear equations. Chapter 3 introduced us to matrices as their own algebraic objects that we could add and multiply by scalars, just as we did for vectors. We learned the matrix–vector product which gave us a compact way to write linear combinations and gave us an efficient way to better analyze systems of linear equations. Finally, in Chapter 4, we introduced spanning sets and linear independence which allowed us to formulate the notion of a subspace and a basis. These concepts relied heavily on the work done in the first three chapters.

As you have likely realized, one does not progress through linear algebra in a linear way. Each new topic we learn ties into many of the previous topics we have already covered, thus giving us new ways to think about past topics. Although this can make learning linear algebra daunting, it also serves to makes linear algebra a rich, fascinating and beautiful subject.

Recall the Matrix Invertibility Criteria. Armed with what we have covered in the current chapter, we now revisit this theorem and add a few new parts. It is the Matrix Invertibility Criteria that truly showcases how interconnected the many topics of linear algebra are.

Theorem 4.7.1 **(Matrix Invertibility Criteria Revisited)**

Let $A \in M_{n\times n}(\mathbb{R})$. The following are equivalent.

(a) A is invertible.

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

(c) The reduced row echelon form of A is I.

(d) For all $\vec{b} \in \mathbb{R}^n$, the system $A\vec{x} = \vec{b}$ is consistent and has a unique solution.

(e) A^T is invertible.

(f) $\operatorname{Null}(A) = \{\vec{0}\}$.

(g) The columns of A form a linearly independent set.

(h) The columns of A span $\mathbb{R}^n$.

(i) The columns of A form a basis for $\mathbb{R}^n$.

(j) $\operatorname{Col}(A) = \mathbb{R}^n$.

Section 4.7 Problems

4.7.1. Prove the following implications in the Matrix Invertibility Criteria Revisited.

(a) $(a) \implies (d)$.

(b) $(a) \implies (f)$.

(c) $(a) \implies (g)$.

(d) $(a) \implies (h)$.

(e) $(a) \implies (i)$.

(f) $(a) \implies (j)$.

4.7.2. Prove the following implications in the Matrix Invertibility Criteria Revisited.

(a) $(d) \implies (a)$.

(b) $(f) \implies (a)$.

(c) $(g) \implies (a)$.

(d) $(h) \implies (a)$.

(e) $(i) \implies (a)$.

(f) $(j) \implies (a)$.

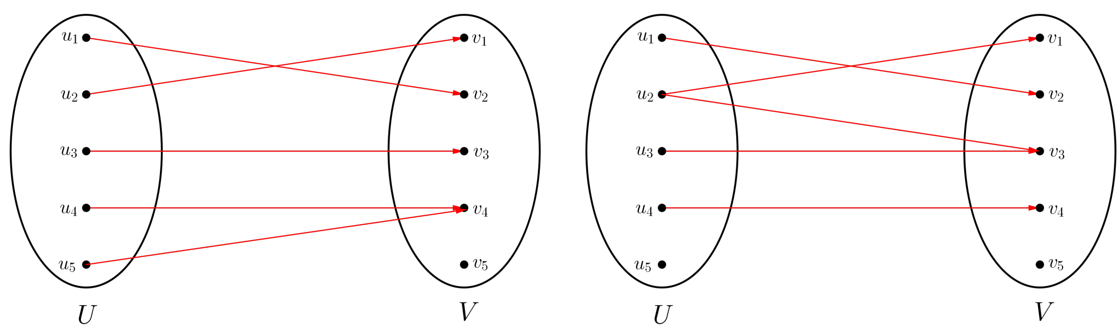
Hint: The implication $(b) \implies (a)$ is proved implicitly in an earlier part of these notes (where?). So instead of proving that $(d), \ldots, (i)$ imply (a) directly, show that they imply (b).

Chapter 5

Linear Transformations

5.1 Matrix Transformations and Linear Transformations

Recall that a *function* is a rule that assigns to every element in one set (called the *domain* of the function) a unique element in another set (called the *codomain*[1] of the function). Given sets U and V we write $f : U \to V$ to indicate that f is a function with domain U and codomain V, and it is understood that to each element $u \in U$, the function f assigns a unique element $v \in V$. We say that f *maps* u to v and that v is the *image* of u under f. We typically write $v = f(u)$. See Figure 5.1.1.



(a) A function with domain U and codomain V.

(b) This fails to be a function from U to V for two reasons: u_5 does not have an image in V, and u_2 has two distinct images in V.

Figure 5.1.1: An example of a function (on the left) and something that fails to be a function (on the right).

In calculus, one studies functions $f : \mathbb{R} \to \mathbb{R}$, for example $f(x) = x^2$ or $f(x) = \sin(x)$. In this course, we will instead consider functions $f : \mathbb{R}^n \to \mathbb{R}^m$. In fact, for $A \in M_{m\times n}(\mathbb{R})$ and $\vec{x} \in \mathbb{R}^n$, we have seen how to compute the matrix–vector product $A\vec{x}$, and we know that $A\vec{x} \in \mathbb{R}^m$. This motivates the following definition.

[1] The codomain of a function is often confused with the *range* of a function. These are different things. We will define the range of a function shortly.

Definition 5.1.1 **Matrix Transformation**

For $A \in M_{m\times n}(\mathbb{R})$, the function $f_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f_A(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$ is called the **matrix transformation corresponding to** A. We call $\mathbb{R}^n$ the **domain** of f_A and $\mathbb{R}^m$ the **codomain** of f_A. We say that f_A **maps** $\vec{x}$ **to** $A\vec{x}$ and say that $A\vec{x}$ is the **image of** $\vec{x}$ **under** f_A.

We make a few notes here:

- It is not uncommon to say *matrix mapping* instead of *matrix transformation*. We may use the words transformation and mapping interchangeably.
- The subscript A in f_A is merely to indicate that the function depends on the matrix A. If we change the matrix A, we change the function f_A.
- For $A \in M_{\mathbf{m}\times\mathbf{n}}(\mathbb{R})$, we have that $f_A : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}^{\mathbf{m}}$. This is a result of how we defined the matrix–vector product.

Example 5.1.2

Let
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 1 \end{bmatrix}.$$
Then $A \in M_{2\times 3}(\mathbb{R})$ and so $f_A : \mathbb{R}^3 \to \mathbb{R}^2$ (that is, the domain of f_A is $\mathbb{R}^3$ and the codomain is $\mathbb{R}^2$). We can compute
$$f_A\left(\begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 4 \end{bmatrix},$$
and more generally,
$$f_A\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ x_1 - x_2 + x_3 \end{bmatrix}.$$

Exercise 74

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

(a) What are the domain and codomain of f_A?

(b) Determine $f_A\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right)$.

Matrix transformations are very special. The next result highlights their two most important algebraic properties.

Theorem 5.1.3 **(Properties of Matrix Transformations)**

Let $A \in M_{m\times n}(\mathbb{R})$ and let f_A be the matrix transformation corresponding to A. For every $\vec{x}, \vec{y} \in \mathbb{R}^n$ and for every $c \in \mathbb{R}$,

(a) $f_A(\vec{x}+\vec{y}) = f_A(\vec{x}) + f_A(\vec{y})$

(b) $f_A(c\vec{x}) = cf_A(\vec{x})$.

Proof: We use the properties of the matrix–vector product as stated in Theorem 3.2.8. We have

$$f_A(\vec{x}+\vec{y}) = A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y} = f_A(\vec{x}) + f_A(\vec{y})$$

and

$$f_A(c\vec{x}) = A(c\vec{x}) = cA\vec{x} = cf_A(\vec{x}). \qquad \square$$

Thus matrix transformations preserve vector sums and scalar multiplication. Combining these two results shows that matrix transformations preserve linear combinations: for $\vec{x}_1, \dots, \vec{x}_k \in \mathbb{R}^n$ and $c_1, \dots, c_k \in \mathbb{R}$,

$$f_A(c_1\vec{x}_1 + \cdots + c_k\vec{x}_k) = c_1 f_A(\vec{x}_1) + \cdots + c_k f_A(\vec{x}_k).$$

Functions which preserve linear combinations are called linear transformations or linear mappings.

Definition 5.1.4
Linear Transformation

A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a **linear transformation** (or a **linear mapping**) if for every $\vec{x}, \vec{y} \in \mathbb{R}^n$ and for every $c \in \mathbb{R}$, we have

T1. $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$ — linear transformations preserve sums

T2. $T(c\vec{x}) = cT(\vec{x})$ — linear transformations preserve scalar multiplication

It follows immediately from Theorem 5.1.3 that every matrix transformation is a linear transformation.

By taking $c = 0$ in T2 of Definition 5.1.4 we find that

$$T(\vec{0}_{\mathbb{R}^n}) = \vec{0}_{\mathbb{R}^m},$$

that is, a linear transformation always sends the zero vector of the domain to the zero vector of the codomain. By taking $c = -1$ in T2, we see that

$$T(-\vec{x}) = -T(\vec{x})$$

so linear transformations preserve negatives as well.

It will become tedious to individually verify T1 and T2 every time we wish to show that a function T is linear. The next theorem presents a more concise way to verify this.

Theorem 5.1.5 **(Linearity Test)**

A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if

$$T(c_1\vec{x} + c_2\vec{y}) = c_1T(\vec{x}) + c_2T(\vec{y})$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and for all $c_1, c_2 \in \mathbb{R}$.

Proof: First assume that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then for $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$,

$$\begin{aligned} T(c_1\vec{x} + c_2\vec{y}) &= T(c_1\vec{x}) + T(c_2\vec{y}) && \text{by T1} \\ &= c_1T(\vec{x}) + c_2T(\vec{y}) && \text{by T2.} \end{aligned}$$

Now assume that the function T satisfies

$$T(c_1\vec{x} + c_2\vec{y}) = c_1T(\vec{x}) + c_2T(\vec{y}) \tag{5.1}$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and for all $c_1, c_2 \in \mathbb{R}$. Since (5.1) holds *for all* $c_1, c_2 \in \mathbb{R}$, we are free to pick any values we like. In particular, substituting $c_1 = c_2 = 1$ in (5.1) gives that

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ so that T1 holds. Taking $c_2 = 0$ in (5.1) gives that

$$T(c_1\vec{x}) = c_1T(\vec{x})$$

for all $\vec{x} \in \mathbb{R}^n$ and all $c_1 \in \mathbb{R}$ so that T2 holds. Thus T is a linear transformation. □

Note that for $\vec{x}_1, \ldots, \vec{x}_k \in \mathbb{R}^n$ and $c_1, \ldots, c_k \in \mathbb{R}$, repeated applications of T1 and T2 show that for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$,

$$T(c_1\vec{x}_1 + \cdots + c_k\vec{x}_k) = c_1T(\vec{x}_1) + \cdots + c_kT(\vec{x}_k)$$

from which we see that linear transformations indeed preserve linear combinations.

Linear transformations are important throughout mathematics – in fact, we have encountered them in calculus.[2] For differentiable functions $f, g : \mathbb{R} \to \mathbb{R}$, and $s, t \in \mathbb{R}$ we have

$$\frac{d}{dx}\left(sf(x) + tg(x)\right) = s\frac{d}{dx}f(x) + t\frac{d}{dx}g(x).$$

This shows that *differentiation is linear.*

Example 5.1.6 Show that $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ 2x_1 + x_2 \end{bmatrix}$$

is a linear transformation.

[2] It is important to always remember that linear algebra is far better than calculus.

Solution: Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

be vectors in $\mathbb{R}^2$, and let $c_1, c_2 \in R$. Then

$$\begin{aligned} T(c_1\vec{x} + c_2\vec{y}) &= T\left(\begin{bmatrix} c_1x_1 + c_2y_1 \\ c_1x_2 + c_2y_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} (c_1x_1 + c_2y_1) - (c_1x_2 + c_2y_2) \\ 2(c_1x_1 + c_2y_1) + (c_1x_2 + c_2y_2) \end{bmatrix} \\ &= \begin{bmatrix} c_1x_1 - c_1x_2 \\ 2c_1x_1 + c_1x_2 \end{bmatrix} + \begin{bmatrix} c_2y_1 - c_2y_2 \\ 2c_2y_1 + c_2y_2 \end{bmatrix} \\ &= c_1\begin{bmatrix} x_1 - x_2 \\ 2x_1 + x_2 \end{bmatrix} + c_2\begin{bmatrix} y_1 - y_2 \\ 2y_1 + y_2 \end{bmatrix} \\ &= c_1T(\vec{x}) + c_2T(\vec{y}). \end{aligned}$$

Thus T is linear by the Linearity Test.

Note that in Example 5.1.6, we could have also observed that for any $\vec{x} \in \mathbb{R}^2$

$$T(\vec{x}) = \begin{bmatrix} x_1 - x_2 \\ 2x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

which shows that T is a matrix transformation and hence a linear transformation.

To show that a function $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is *not* a linear transformation, it is sufficient to show that at least one of T1 or T2 fail to hold, that is, it is enough to show that either

- T1 fails, that is, there exist two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$ so that $T(\vec{x} + \vec{y}) \neq T(\vec{x}) + T(\vec{y})$
- T2 fails, that is, there exists a vector $\vec{x} \in \mathbb{R}^n$ and scalar $c \in \mathbb{R}$ so that $T(c\vec{x}) \neq cT(\vec{x})$.

Let's look at a couple of examples.

Example 5.1.7 Show that $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_3^2 + 3 \end{bmatrix}$$

is not linear.

Solution: Consider $\vec{x} = \left[\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right]$ and $\vec{y} = \left[\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right]$. Then

$$T(\vec{x} + \vec{y}) = T\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

but

$$T(\vec{x})+T(\vec{y}) = T\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right)+T\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}1\\3\end{bmatrix}+\begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}2\\6\end{bmatrix}.$$

Since $T(\vec{x})+T(\vec{y}) \neq T(\vec{x})+T(\vec{y})$ (that is T does not preserve sums), T is not linear.

Exercise 75 Let T be defined as in Example 5.1.7. Find $\vec{x} \in \mathbb{R}^3$ and $c \in \mathbb{R}$ so that $T(c\vec{x}) \neq cT(\vec{x})$, that is, show that T does not preserve scalar multiplication.

Recall that a linear transformation always maps the zero vector of the domain to the zero vector of the codomain. Thus in Example 5.1.7, we could have quickly noticed that

$$T\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\right) = \begin{bmatrix}0\\3\end{bmatrix} \neq \begin{bmatrix}0\\0\end{bmatrix}$$

and concluded immediately that T was not linear. Note however, that a function sending the zero vector of the domain to the zero vector of the codomain does not guarantee that the function is linear as is illustrated in the next example.

Example 5.1.8 Show that $T : \mathbb{R}^2 \to \mathbb{R}$ defined by $T(\vec{x}) = \|\vec{x}\|$ is not linear.

Solution: Let $\vec{x} = \vec{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $c = -1$. Then

$$T(c\vec{x}) = T(-\vec{e}_1) = \|-\vec{e}_1\| = |-1|\|\vec{e}_1\| = 1,$$

but

$$cT(\vec{x}) = -T(\vec{e}_1) = -\|\vec{e}_1\| = -1.$$

Since $T(c\vec{x}) \neq cT(\vec{x})$ (that is, T does not preserve scalar multiplication), T is not linear.

Exercise 76 Let T be defined as in Example 5.1.8. Find $\vec{x}, \vec{y} \in \mathbb{R}^3$ so that $T(\vec{x}+\vec{y}) \neq T(\vec{x})+T(\vec{y})$, that is, show that T does not preserve sums.

The next example shows a very important and very useful property of linear transformations.

Example 5.1.9 Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation such that

$$T(\vec{e}_1) = \begin{bmatrix}1\\-2\\1\end{bmatrix} \quad \text{and} \quad T(\vec{e}_2) = \begin{bmatrix}0\\1\\-1\end{bmatrix}$$

(a) Compute $T\left(\begin{bmatrix}3\\5\end{bmatrix}\right)$.

(b) Compute $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$.

Solution:

(a) Since $\left[\begin{smallmatrix} 3 \\ 5 \end{smallmatrix}\right] = 3\vec{e}_1 + 5\vec{e}_2$, we use that fact that linear transformations preserve linear combinations to compute

$$T\left(\begin{bmatrix} 3 \\ 5 \end{bmatrix}\right) = T(3\vec{e}_1 + 5\vec{e}_2) = 3T(\vec{e}_1) + 5T(\vec{e}_2) = 3\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}.$$

(b) Proceeding as in part (a) with $\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] = x_1\vec{e}_1 + x_2\vec{e}_2$, we have

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = T(x_1\vec{e}_1 + x_2\vec{e}_2) = x_1T(\vec{e}_1) + x_2T(\vec{e}_2) = x_1\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} x_1 \\ -2x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}.$$

In Example 5.1.9, we were able to compute $T(\vec{x})$ for *any* $\vec{x} \in \mathbb{R}^2$ knowing *only* $T(\vec{e}_1)$ and $T(\vec{e}_2)$. In general, for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, if we are given $T(\vec{x}_1), \dots, T(\vec{x}_k)$ for $\vec{x}_1, \dots, \vec{x}_k \in \mathbb{R}^n$, then we can compute $T(\vec{x})$ for any $\vec{x} \in \operatorname{Span}\{\vec{x}_1, \dots, \vec{x}_k\}$ since T preserves linear combinations. In particular, if $\{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis for $\mathbb{R}^n$ and we know $T(\vec{v}_1), \dots, T(\vec{v}_n)$, then we can compute $T(\vec{v})$ for any $\vec{v} \in \mathbb{R}^n$ which is an extremely powerful property!

It is also worth noting that in Example 5.1.9 with $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$,

$$T(\vec{x}) = \begin{bmatrix} x_1 \\ -2x_1 + x_2 \\ x_1 - x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_2 \\ x_2 \end{bmatrix}$$

which shows that T is a matrix transformation. We also saw that the linear transformation in Example 5.1.6 is a matrix transformation, too. This is not a coincidence! The next result shows that *every* linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation.

Theorem 5.1.10 If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then T is a matrix transformation with corresponding matrix

$$\big[\,T\,\big] = \big[\,T(\vec{e}_1) \ \ \cdots \ \ T(\vec{e}_n)\,\big] \in M_{m\times n}(\mathbb{R}),$$

that is, $T(\vec{x}) = \big[\,T\,\big]\,\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$.

It is important to note that the matrix $\big[\,T(\vec{e}_1) \ \ \cdots \ \ T(\vec{e}_n)\,\big]$ is the $m \times n$ matrix whose columns are $T(\vec{e}_1), \dots, T(\vec{e}_n)$. Thus, in order to construct $\big[\,T\,\big]$, we need only compute $T(\vec{e}_1), \dots, T(\vec{e}_n)$.

Proof (of Theorem 5.1.10): Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$. Then $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$. We have

$$\begin{aligned} T(\vec{x}) &= T(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) \\ &= x_1T(\vec{e}_1) + \cdots + x_nT(\vec{e}_n) \quad \text{since } T \text{ is linear} \\ &= \begin{bmatrix} T(\vec{e}_1) & \cdots & T(\vec{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} T \end{bmatrix} \vec{x}. \end{aligned}$$

□

Note that Theorem 5.1.3 and Theorem 5.1.10 combine to give that T is linear if and only if it is a matrix transformation. This motivates the following definition.

Definition 5.1.11
Standard Matrix

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The matrix

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} T(\vec{e}_1) & \cdots & T(\vec{e}_n) \end{bmatrix} \in M_{m\times n}(\mathbb{R})$$

is called the **standard matrix** of T.

Example 5.1.12

Returning to the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ 2x_1 + x_2 \end{bmatrix}$$

given in Example 5.1.6, we have

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) \end{bmatrix} = \left[T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \; T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \right] = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}.$$

Notice that

$$\begin{bmatrix} T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ 2x_1 + x_2 \end{bmatrix}$$

which shows that $\begin{bmatrix} T \end{bmatrix} \vec{x} = T(\vec{x})$. So the linear transformation T is equal to the matrix transformation defined by $\begin{bmatrix} T \end{bmatrix}$.

Example 5.1.13

Let $T \colon \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

That is, $T(\vec{x}) = \vec{x}$ for all $\vec{x} \in \mathbb{R}^3$. (We call T the *identity transformation on* $\mathbb{R}^3$.) Find the standard matrix of T.

Solution: We have

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) & T(\vec{e}_3) \end{bmatrix} = \left[T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) \; T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) \; T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) \right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

That is, the standard matrix of the identity transformation is the identity matrix! Of course, this makes sense since

$$T(\vec{x}) = \vec{x} = I\vec{x} = \left[T\right]\vec{x}.$$

Example 5.1.13 is actually quite important and is worthy of a definition. We will encounter the identity transformation again in Section 5.4.

Definition 5.1.14 **Identity Transformation**

The linear transformation $\mathrm{Id}_n : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\mathrm{Id}_n(\vec{x}) = \vec{x}$ for every $\vec{x} \in \mathbb{R}^n$ is called the **identity transformation**.

Computing the standard matrix for $\mathrm{Id}_n : \mathbb{R}^n \to \mathbb{R}^n$ gives

$$\left[\mathrm{Id}_n\right] = \left[\mathrm{Id}_n(\vec{e}_1) \ \cdots \ \mathrm{Id}_n(\vec{e}_n)\right] = \left[\vec{e}_1 \ \cdots \ \vec{e}_n\right] = I_n.$$

That is, the standard matrix of the Id_n is the $n \times n$ identity matrix I_n.

Note that just as we write I instead of I_n when there is no confusion, we may write Id instead of Id_n.

Exercise 77

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \\ 2x_1 + 3x_2 \end{bmatrix}.$$

Determine the standard matrix of T.

The next example is a little trickier.

Example 5.1.15

Let $T : \mathbb{R}^2 \to \mathbb{R}^4$ be a linear transformation such that

$$T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 4 \\ 0 \\ -1 \end{bmatrix}.$$

Determine the standard matrix of T.

Solution: Our goal is to compute $\left[T\right]$, so we need to compute $T(\vec{e}_1)$ and $T(\vec{e}_2)$. We will be able to do this if we can express $\vec{e}_1$ and $\vec{e}_2$ as linear combinations of $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$, since for instance if

$$\vec{e}_1 = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

then by the linearity of T,

$$T(\vec{e}_1) = c_1 T\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) + c_2 T\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = c_1 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 4 \\ 0 \\ -1 \end{bmatrix},$$

and similarly for $T(\vec{e}_2)$. Notice that since $\{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right],\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]\}$ is linearly independent, $\{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right],\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]\}$ is a basis for $\mathbb{R}^2$, and so we are guaranteed that $\vec{e}_1, \vec{e}_2 \in \operatorname{Span}\{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right],\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]\}$. Hence we can express $\vec{e}_1$ and $\vec{e}_2$ as linear combinations of $\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]$. Since

$$\left[\begin{array}{cc|cc}1&2&1&0\\2&3&0&1\end{array}\right] \xrightarrow[R_2-2R_1]{} \left[\begin{array}{cc|cc}1&2&1&0\\0&-1&-2&1\end{array}\right] \xrightarrow[-R_2]{} \left[\begin{array}{cc|cc}1&2&1&0\\0&1&2&-1\end{array}\right] \xrightarrow{R_1-2R_2} \left[\begin{array}{cc|cc}1&0&-3&2\\0&1&2&-1\end{array}\right]$$

we have that

$$\vec{e}_1 = -3\begin{bmatrix}1\\2\end{bmatrix} + 2\begin{bmatrix}2\\3\end{bmatrix} \quad \text{and} \quad \vec{e}_2 = 2\begin{bmatrix}1\\2\end{bmatrix} - 1\begin{bmatrix}2\\3\end{bmatrix}.$$

Thus

$$T(\vec{e}_1) = T\left(-3\begin{bmatrix}1\\2\end{bmatrix} + 2\begin{bmatrix}2\\3\end{bmatrix}\right) = -3T\left(\begin{bmatrix}1\\2\end{bmatrix}\right) + 2T\left(\begin{bmatrix}2\\3\end{bmatrix}\right) = -3\begin{bmatrix}1\\2\\3\\4\end{bmatrix} + 2\begin{bmatrix}1\\4\\0\\-1\end{bmatrix} = \begin{bmatrix}-1\\2\\-9\\-14\end{bmatrix}$$

$$T(\vec{e}_2) = T\left(2\begin{bmatrix}1\\2\end{bmatrix} - 1\begin{bmatrix}2\\3\end{bmatrix}\right) = 2T\left(\begin{bmatrix}1\\2\end{bmatrix}\right) - 1T\left(\begin{bmatrix}2\\3\end{bmatrix}\right) = 2\begin{bmatrix}1\\2\\3\\4\end{bmatrix} - 1\begin{bmatrix}1\\4\\0\\-1\end{bmatrix} = \begin{bmatrix}1\\0\\6\\9\end{bmatrix}.$$

Hence

$$[T] = [T(\vec{e}_1) \;\; T(\vec{e}_2)] = \begin{bmatrix}-1&1\\2&0\\-9&6\\-14&9\end{bmatrix}.$$

Exercise 78 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation such that

$$T\left(\begin{bmatrix}1\\0\\1\end{bmatrix}\right) = \begin{bmatrix}2\\4\\1\end{bmatrix}, \quad T\left(\begin{bmatrix}1\\0\\-1\end{bmatrix}\right) = \begin{bmatrix}0\\2\\1\end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix}0\\1\\1\end{bmatrix}\right) = \begin{bmatrix}-1\\1\\3\end{bmatrix}.$$

Find the standard matrix of T.

Section 5.1 Problems

5.1.1. Consider the linear transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 - 2x_3 \\ 3x_1 + 6x_3 \end{bmatrix}.$$

(a) State the domain of T.

(b) State the codomain of T.

(c) Evaluate $T\left(\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right]\right)$.

(d) Find all $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right] \in \mathbb{R}^3$ so that $T(\vec{x}) = \left[\begin{smallmatrix} -3 \\ 21 \end{smallmatrix}\right]$.

5.1.2. For each of the following, either show T is a linear transformation using the Linearity Test, or give an example to show that T is not a linear transformation.

(a) $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2 \\ x_1 - x_2 \\ 2x_2 \end{bmatrix}$.

(b) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 + 2x_3 \\ 0 \\ 2x_1 - 3x_3 \end{bmatrix}$.

5.1.3. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ be such that

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Determine whether or not T is a linear transformation.

5.1.4. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation such that

$$T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 4 \end{bmatrix}.$$

(a) Compute $\big[\, T \,\big]$, the standard matrix of T.

(b) Use $\big[\, T \,\big]$ to find an expression for $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$.

5.1.5. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ be a linear transformation satisfying

$$T\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad T\left(\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

Determine $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)$.

5.1.6. Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$. Prove that if $T(\vec{v}_1) = \cdots = T(\vec{v}_n) = \vec{0}$, then $T(\vec{x}) = \vec{0}$ for all $\vec{x} \in \mathbb{R}^n$.

5.1.7. Let $A \in M_{m\times n}(\mathbb{R})$ and let $f_A\colon \mathbb{R}^n \to \mathbb{R}^m$ be the associated matrix transformation. Since f_A is a linear transformation, it has a standard matrix $\big[\, f_A \,\big]$. Determine $\big[\, f_A \,\big]$.

5.2 Examples of Linear Transformations

Having defined linear transformations and stated some of their important properties, we turn our attention to looking at some meaningful examples. We will see that many common geometric transformations can be represented by linear transformations. Of course, since every linear transformation is a matrix transformation, we will, at the same time, gain a geometric interpretation of the matrix–vector product.

We first show that projections are linear transformations.

Theorem 5.2.1 Let $\vec{d} \in \mathbb{R}^n$ with $\vec{d} \neq \vec{0}$.

(a) The function $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(\vec{x}) = \operatorname{proj}_{\vec{d}} \vec{x}$ is linear.

(b) The function $S : \mathbb{R}^n \to \mathbb{R}^n$ defined by $S(\vec{x}) = \operatorname{perp}_{\vec{d}} \vec{x}$ is linear.

Proof: We prove part (a). Let $\vec{x}, \vec{y} \in \mathbb{R}^2$ and $c_1, c_2 \in \mathbb{R}$. Then

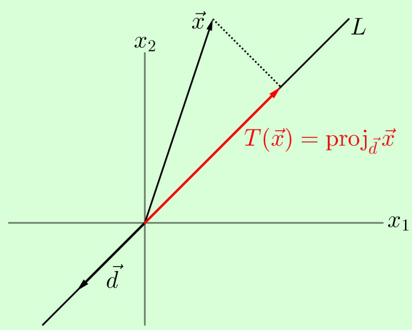
$$\begin{aligned}
T(c_1\vec{x} + c_2\vec{y}) &= \operatorname{proj}_{\vec{d}}(c_1\vec{x} + c_2\vec{y}) \\
&= \frac{(c_1\vec{x} + c_2\vec{y}) \cdot \vec{d}}{\|\vec{d}\|^2}\vec{d} \\
&= \frac{(c_1\vec{x}) \cdot \vec{d}}{\|\vec{d}\|^2}\vec{d} + \frac{(c_2\vec{y}) \cdot \vec{d}}{\|\vec{d}\|^2}\vec{d} \\
&= c_1\frac{\vec{x} \cdot \vec{d}}{\|\vec{d}\|^2}\vec{d} + c_2\frac{\vec{y} \cdot \vec{d}}{\|\vec{d}\|^2}\vec{d} \\
&= c_1 \operatorname{proj}_{\vec{d}} \vec{x} + c_2 \operatorname{proj}_{\vec{d}} \vec{y} \\
&= c_1T(\vec{x}) + c_2T(\vec{y}),
\end{aligned}$$

and thus T is linear by the Linearity Test. □

Exercise 79 Prove Theorem 5.2.1(b).

Example 5.2.2 **(Projection Onto a Line in $\mathbb{R}^2$)**

Let $\vec{d} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] \in \mathbb{R}^2$ and define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\vec{x}) = \operatorname{proj}_{\vec{d}} \vec{x}$. The following figure shows that we may view T as a projection onto L where L is the line through the origin with direction vector $\vec{d}$.



Find the standard matrix of T.

Solution: It follows from Theorem 5.2.1(a) that T is a linear transformation. By definition, the standard matrix of T is $\big[T\big] = \big[T(\vec{e}_1) \;\; T(\vec{e}_2)\big]$. We have

$$T(\vec{e}_1) = \operatorname{proj}_{\vec{d}}\, \vec{e}_1 = \frac{\vec{e}_1 \cdot \vec{d}}{\|\vec{d}\|^2}\vec{d} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1/2\\1/2\end{bmatrix}$$
$$T(\vec{e}_2) = \operatorname{proj}_{\vec{d}}\, \vec{e}_2 = \frac{\vec{e}_2 \cdot \vec{d}}{\|\vec{d}\|^2}\vec{d} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1/2\\1/2\end{bmatrix}$$

so

$$\big[T\big] = \big[T(\vec{e}_1) \;\; T(\vec{e}_2)\big] = \begin{bmatrix}1/2 & 1/2\\1/2 & 1/2\end{bmatrix}.$$

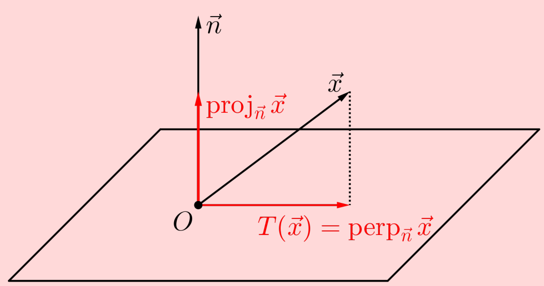
Note that if we take $\vec{x} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$, for example, we can compute the projection of $\vec{x}$ onto $\vec{d} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ as

$$\operatorname{proj}_{\vec{d}}\, \vec{x} = T(\vec{x}) = \begin{bmatrix}1/2 & 1/2\\1/2 & 1/2\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}3/2\\3/2\end{bmatrix},$$

that is, we can compute projections using the matrix–vector product!

Exercise 80 **(Projection onto a Plane in $\mathbb{R}^3$)**

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(\vec{x}) = \operatorname{perp}_{\vec{n}}\, \vec{x}$ for all $\vec{x} \in \mathbb{R}^3$ where $\vec{n} \in \mathbb{R}^3$ is a nonzero vector. The figure below shows that T represents a *projection* onto the plane through the origin with normal vector $\vec{n}$, that is, if P and Q are such that $\overrightarrow{OP} = \vec{x}$ and $\overrightarrow{OQ} = T(\vec{x})$, then Q is the closest point on the plane to P.

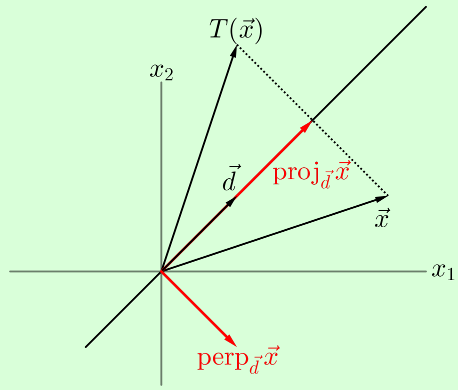


Find the standard matrix for T with $\vec{n} = \left[\begin{smallmatrix}2\\-1\\1\end{smallmatrix}\right]$.

We now look at how projections can be used to define reflections, which form another important class of geometric transformations.

Example 5.2.3 **(Reflection Through a Line in $\mathbb{R}^2$)**

Let $\vec{d} \in \mathbb{R}^2$ be a nonzero vector, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(\vec{x}) = \vec{x} - 2\operatorname{perp}_{\vec{d}}\, \vec{x}$ for every $\vec{x} \in \mathbb{R}^2$. The figure below shows that T represents a reflection through the line containing the origin with direction vector $\vec{d}$.



Note that

$$T(\vec{x}) = \vec{x} - 2\,\text{perp}_{\vec{d}}\,\vec{x} = \vec{x} - 2(\vec{x} - \text{proj}_{\vec{d}}\,\vec{x}) = 2\,\text{proj}_{\vec{d}}\,\vec{x} - \vec{x}$$

and we will prefer to work with $T(\vec{x}) = 2\,\text{proj}_{\vec{d}}\,\vec{x} - \vec{x}$. Show that T is linear, and then find the standard matrix of T with $\vec{d} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$.

Solution: We first show that T is linear. For $\vec{x}, \vec{y} \in \mathbb{R}^2$ and $c_1, c_2 \in \mathbb{R}$, we have

$$\begin{aligned} T(c_1\vec{x} + c_2\vec{y}) &= 2\,\text{proj}_{\vec{d}}(c_1\vec{x} + c_2\vec{y}) - (c_1\vec{x} + c_2\vec{y}) \\ &= 2(c_1\,\text{proj}_{\vec{d}}\,\vec{x} + c_2\,\text{proj}_{\vec{d}}\,\vec{y}) - c_1\vec{x} - c_2\vec{y} \qquad \text{by Theorem 5.2.1(a)} \\ &= c_1(2\,\text{proj}_{\vec{d}}\,\vec{x} - \vec{x}) + c_2(2\,\text{proj}_{\vec{d}}\,\vec{y} - \vec{y}) \\ &= c_1T(\vec{x}) + c_2T(\vec{y}). \end{aligned}$$

Thus, by the Linearity Test, T is linear. Now with $\vec{d} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$,

$$\begin{aligned} T(\vec{e}_1) &= 2\,\text{proj}_{\vec{d}}\,\vec{e}_1 - \vec{e}_1 = 2\left(\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}\right) - \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix} \\ T(\vec{e}_2) &= 2\,\text{proj}_{\vec{d}}\,\vec{e}_2 - \vec{e}_2 = 2\left(\frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}\right) - \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix} \end{aligned}$$

and so

$$\begin{bmatrix}T\end{bmatrix} = \begin{bmatrix}T(\vec{e}_1) & T(\vec{e}_2)\end{bmatrix} = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}.$$

Note that in Example 5.2.3, for any $\vec{x} = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right] \in \mathbb{R}^2$,

$$T(\vec{x}) = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}x_2\\x_1\end{bmatrix}$$

from which we see that reflecting a vector in the line with direction vector $\vec{d} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ simply swaps the coordinates of that vector.

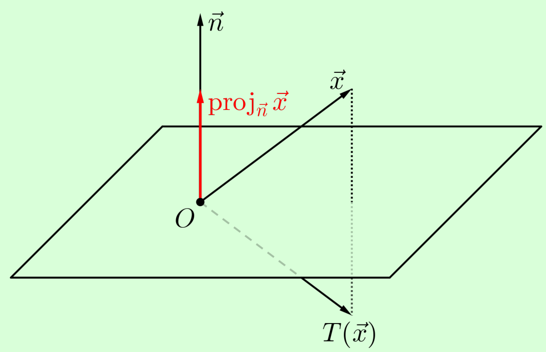
Example 5.2.4 **(Reflection Through a Plane in $\mathbb{R}^3$)**

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(\vec{x}) = \vec{x} - 2\operatorname{proj}_{\vec{n}} \vec{x}$ for all $\vec{x} \in \mathbb{R}^3$ where $\vec{n} \in \mathbb{R}^3$ is a nonzero vector. The figure below shows that T represents a reflection through the plane containing the origin with normal vector $\vec{n}$.

(a) Show that T is linear.

(b) Find the standard matrix of T if the plane has scalar equation $x_1 - x_2 + 2x_3 = 0$.

(c) Find the vector $\vec{y} \in \mathbb{R}^3$ that is the result of reflecting the vector $\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ through the plane with scalar equation $x_1 - x_2 + 2x_3 = 0$.



Solution: (a) For $\vec{x}, \vec{y} \in \mathbb{R}^3$, and $c_1, c_2 \in \mathbb{R}$,

$$\begin{aligned} T(c_1\vec{x} + c_2\vec{y}) &= (c_1\vec{x} + c_2\vec{y}) - 2\operatorname{proj}_{\vec{n}}(c_1\vec{x} + c_2\vec{y}) \\ &= c_1\vec{x} + c_2\vec{y} - 2(c_1 \operatorname{proj}_{\vec{n}} \vec{x} + c_2 \operatorname{proj}_{\vec{n}} \vec{y}) \qquad \text{by Theorem 5.2.1(a)} \\ &= c_1(\vec{x} - 2\operatorname{proj}_{\vec{n}} \vec{x}) + c_2(\vec{y} - 2\operatorname{proj}_{\vec{n}} \vec{y}) \\ &= c_1T(\vec{x}) + c_2T(\vec{y}) \end{aligned}$$

and so T is linear.

(b) Now for the plane $x_1 - x_2 + 2x_3 = 0$, we have that $\vec{n} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$. We compute

$$T(\vec{e}_1) = \vec{e}_1 - 2\operatorname{proj}_{\vec{n}} \vec{e}_1 = \vec{e}_1 - 2\frac{\vec{e}_1 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - 2\left(\frac{1}{6}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 2/3 \\ 1/3 \\ -2/3 \end{bmatrix}$$

$$T(\vec{e}_2) = \vec{e}_2 - 2\operatorname{proj}_{\vec{n}} \vec{e}_2 = \vec{e}_2 - 2\frac{\vec{e}_2 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 2\left(\frac{(-1)}{6}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

$$T(\vec{e}_3) = \vec{e}_3 - 2\operatorname{proj}_{\vec{n}} \vec{e}_3 = \vec{e}_3 - 2\frac{\vec{e}_3 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - 2\left(\frac{2}{6}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -2/3 \\ 2/3 \\ -1/3 \end{bmatrix}$$

Hence the standard matrix of T is

$$\big[\,T\,\big] = \big[\,T(\vec{e}_1) \;\; T(\vec{e}_2) \;\; T(\vec{e}_3)\,\big] = \begin{bmatrix} 2/3 & 1/3 & -2/3 \\ 1/3 & 2/3 & 2/3 \\ -2/3 & 2/3 & -1/3 \end{bmatrix}.$$

(c) We wish to determine $\vec{y} = T(\vec{x})$. We have

$$\vec{y} = T(\vec{x}) = \begin{bmatrix} T \end{bmatrix} \vec{x} = \begin{bmatrix} 2/3 & 1/3 & -2/3 \\ 1/3 & 2/3 & 2/3 \\ -2/3 & 2/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -1/3 \\ -4/3 \end{bmatrix}.$$

In Example 5.2.3 and Example 5.2.4, we required the "objects" we were reflecting through (lines and planes) to contain the origin. The reason for this is because if our line or plane does not contain the origin, then these transformations would not send the zero vector to the zero vector and thus not be linear.

We are seeing that linear transformations (or equivalently, matrix transformations) give us a way to geometrically understand the matrix–vector product. Having seen that projections and reflections are both linear transformations, we now look at some additional linear transformations that are common in many fields, such as computer graphics.

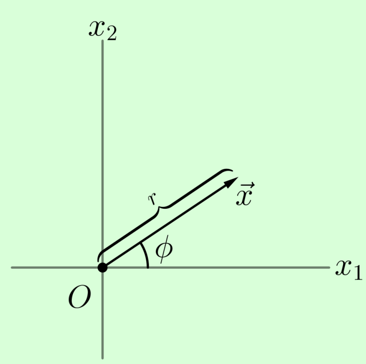
We first consider rotations.

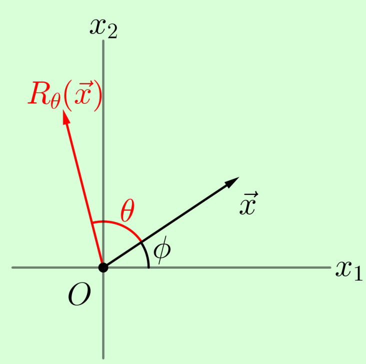
Example 5.2.5 **(Rotations in $\mathbb{R}^2$)**

Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be a counterclockwise rotation about the origin by an angle of θ. To see that R_θ is linear, we use basic trigonometry to write $\vec{x} \in \mathbb{R}^2$ as

$$\vec{x} = \begin{bmatrix} r\cos\phi \\ r\sin\phi \end{bmatrix}$$

where $r \in \mathbb{R}$ satisfies $r = \|\vec{x}\| \geq 0$ and $\phi \in \mathbb{R}$ is the angle $\vec{x}$ makes with the positive x_1-axis measured counterclockwise (if $\vec{x} = \vec{0}$, then $r = 0$ and we may take ϕ to be any real number).





Since $R_\theta(\vec{x})$ is obtained from rotating $\vec{x}$ counterclockwise about the origin, it is clear that $\|R_\theta(\vec{x})\| = r$ and that $R_\theta(\vec{x})$ makes an angle of $\phi + \theta$ with the positive x_1-axis. Thus using the angle-sum formulas for sine and cosine, we have

$$\begin{aligned} R_\theta(\vec{x}) &= \begin{bmatrix} r\cos(\phi+\theta) \\ r\sin(\phi+\theta) \end{bmatrix} \\ &= \begin{bmatrix} r(\cos\phi\cos\theta - \sin\phi\sin\theta) \\ r(\sin\phi\cos\theta + \cos\phi\sin\theta) \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \cos\theta(r\cos\phi) - \sin\theta(r\sin\phi) \\ \sin\theta(r\cos\phi) + \cos\theta(r\sin\phi) \end{bmatrix}$$
$$= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r\cos\phi \\ r\sin\phi \end{bmatrix}$$
$$= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \vec{x}$$

and we see that R_θ is a matrix transformation and thus a linear transformation. We also see that

$$[\,R_\theta\,] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Example 5.2.6 Find the vector that results from rotating $\vec{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ counterclockwise about the origin by an angle of $\frac{\pi}{6}$.

Solution: We have

$$R_{\frac{\pi}{6}}(\vec{x}) = [\,R_{\frac{\pi}{6}}\,]\vec{x} = \begin{bmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sqrt{3}-2 \\ 1+2\sqrt{3} \end{bmatrix}.$$

Note that a clockwise rotation about the origin by an angle of θ is simply a counterclockwise rotation about the origin by an angle of $-\theta$. Thus a clockwise rotation by θ is given by the linear transformation

$$[\,R_{-\theta}\,] = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

where we have used the fact that $\cos\theta$ is an even function and $\sin\theta$ is an odd function, that is,

$$\cos(-\theta) = \cos\theta \quad \text{and} \quad \sin(-\theta) = -\sin\theta.$$

We briefly mention that we can generalize these results for rotations about a coordinate axis in $\mathbb{R}^3$. Consider[3]

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}, \quad B = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, \quad C = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then

$T_1 : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T_1(\vec{x}) = A\vec{x}$ is a counterclockwise rotation about the x_1- axis,
$T_2 : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T_2(\vec{x}) = B\vec{x}$ is a counterclockwise rotation about the x_2- axis,
$T_3 : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T_3(\vec{x}) = C\vec{x}$ is a counterclockwise rotation about the x_3- axis.

[3] For the matrix B, notice that the negative sign is on the "other" instance of $\sin\theta$. The reason for this is if one "stares" down the positive x_2-axis towards the origin, then one sees the x_1x_3-plane, however, the orientation is backwards – the positive x_1-axis is to the left of the positive x_3-axis. Thus the roles of "clockwise" and "counterclockwise" are reversed in this instance.

In fact, we can rotate about any line through the origin in $\mathbb{R}^3$, but finding the standard matrix of such a transformation is beyond the scope of this course.

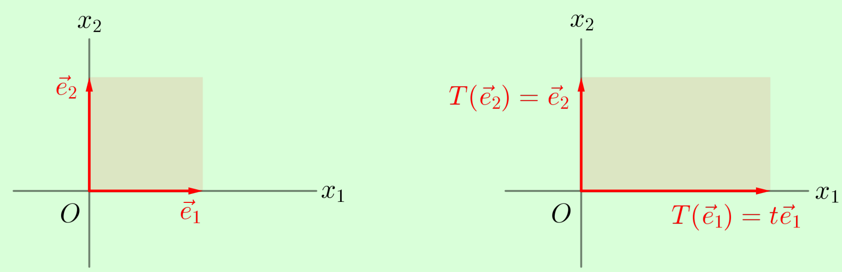
We next look at stretches and compressions.

Example 5.2.7 **(Stretches and Compressions in $\mathbb{R}^2$)**

For $t \in \mathbb{R}$ with $t > 0$, let
$$A = \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix}$$
and define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\vec{x}) = A\vec{x}$ for every $\vec{x} \in \mathbb{R}^2$. Then T is a matrix transformation and hence a linear transformation. For $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$,
$$T(\vec{x}) = \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} tx_1 \\ x_2 \end{bmatrix}.$$

If $t > 1$, then we say that T is a *stretch* in the x_1-direction by a factor of t (also called a horizontal stretch by a factor of t), and if $0 < t < 1$, we say that T is a *compression* in the x_1-direction by a factor of t (also called a horizontal compression by a factor of t). If $t = 1$, then A is the identity matrix and $T(\vec{x}) = \vec{x}$. A stretch or compression in the x_2-direction is defined in a similar way.

A stretch in the $\vec{x}_1$-direction is illustrated below.



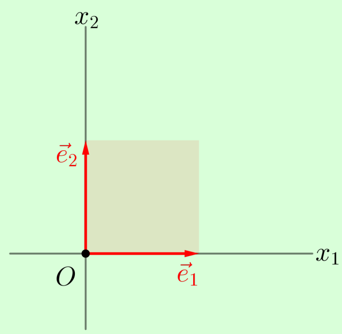
Note the requirement that $t > 0$. If $t = 0$, then T is actually a projection onto the x_2-axis, and if $t < 0$, then T is a reflection in the x_2-axis followed by a stretch or compression by a factor of $-t > 0$.

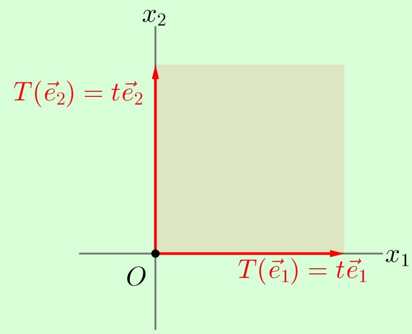
Exercise 81 Write down the standard matrix for a stretch or compression in the x_2-direction by a factor of $t > 0$.

Example 5.2.8 **(Dilations and Contractions in $\mathbb{R}^2$)**

For $t \in \mathbb{R}$ with $t > 0$, let
$$B = \begin{bmatrix} t & 0 \\ 0 & t \end{bmatrix}$$
and define $T(\vec{x}) = B\vec{x}$ for every $\vec{x} \in \mathbb{R}^2$. Then T is a matrix transformation and thus a linear transformation. For $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$,
$$T(\vec{x}) = \begin{bmatrix} t & 0 \\ 0 & t \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} tx_1 \\ tx_2 \end{bmatrix} = t\vec{x}.$$

We see that $T(\vec{x})$ is simply a scalar multiple of $\vec{x}$. We call T a *dilation* by a factor of t if $t > 1$ and we call T a *contraction* by a factor of t if $0 < t < 1$. If $t = 1$, then B is the identity matrix and $T(\vec{x}) = \vec{x}$. A dilation is illustrated below.





Example 5.2.9 **(Horizontal Shear in $\mathbb{R}^2$)**

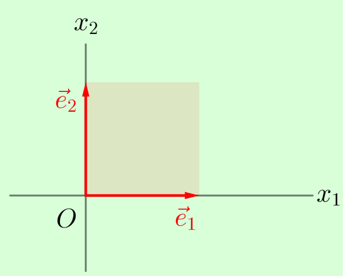
For $s \in \mathbb{R}$, let

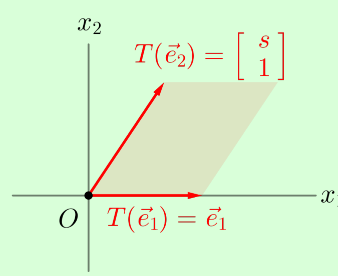
$$C = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$$

and define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\vec{x}) = C\vec{x}$ for every $\vec{x} \in \mathbb{R}^2$. Then T is a matrix transformation and hence a linear transformation. For $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$,

$$T(\vec{x}) = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + sx_2 \\ x_2 \end{bmatrix}.$$

We call T a *shear* in the x_1-direction by a factor of s (also referred to as a horizontal shear by a factor of s). If $s = 0$, then C is the identity matrix and $T(\vec{x}) = \vec{x}$. A shear in the x_1-direction is illustrated below (with $s > 0$).





Exercise 82 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a shear in the x_2-direction by a factor of 3 (also referred to as a vertical shear by a factor of 3). Determine the standard matrix of T and hence find $T\left(\left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]\right)$.

Let's summarize our findings in the table below.

Linear Transformation in $\mathbb{R}^2$	Standard Matrix
Counterclockwise rotation by $\theta \in [0, 2\pi)$	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$
Horizontal stretch/compression by $t > 0$	$\begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix}$
Vertical stretch/compression by $t > 0$	$\begin{bmatrix} 1 & 0 \\ 0 & t \end{bmatrix}$
Dilation/contraction by $t > 0$	$\begin{bmatrix} t & 0 \\ 0 & t \end{bmatrix}$
Horizontal shear by s	$\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$
Vertical shear by s	$\begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}$

Section 5.2 Problems

5.2.1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a projection onto the line that passes through the origin with direction vector $\vec{d} = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$. Determine $\big[T\big]$ and use it to compute $T(\vec{x})$ for $\vec{x} = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$.

5.2.2. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a reflection through the line that passes through the origin with direction vector $\left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$. Determine $\big[T\big]$ and use it to compute $T(\vec{x})$ for $\vec{x} = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$.

5.2.3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a counterclockwise rotation about the origin by an angle of $\frac{\pi}{3}$. Compute $T(\vec{x})$ where $\vec{x} = \left[\begin{smallmatrix} 1 \\ -2 \end{smallmatrix}\right]$.

5.2.4. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a projection onto the plane with scalar equation $\sqrt{2}x_1 + x_2 + x_3 = 0$. Determine $\big[T\big]$ and use it to compute $T(\vec{x})$ for $\vec{x} = \begin{bmatrix} 4 \\ -8 \\ -4 \end{bmatrix}$.

5.2.5. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a reflection through the plane with scalar equation $x_1 - 2x_2 + 2x_3 = 0$. Determine $\big[T\big]$ and use it to compute $T(\vec{x})$ for $\vec{x} = \left[\begin{smallmatrix} 9 \\ 27 \\ 18 \end{smallmatrix}\right]$.

5.2.6. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a horizontal shear such that $T\left(\left[\begin{smallmatrix} 5 \\ 6 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} -7 \\ 6 \end{smallmatrix}\right]$. Determine $\big[T\big]$.

5.2.7. Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ denote a reflection through the line containing the origin with direction vector $\vec{d} = \left[\begin{smallmatrix} d_1 \\ d_2 \end{smallmatrix}\right] \neq \vec{0}$. Determine $\big[R\big]$.

5.2.8. Let $\vec{u} \in \mathbb{R}^3$ be a unit vector and let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(\vec{x}) = \operatorname{proj}_{\vec{u}} \vec{x}$. Show that $\big[T\big] = \vec{u}\,\vec{u}^T$.

5.3 Operations on Linear Transformations

We now study linear transformations more algebraically. Given the relationship between linear transformations and matrices, it shouldn't be too much of a surprise that we obtain similar results for linear transformations as we did for matrices in Chapter 3.

Definition 5.3.1 **Zero Transformation**

The function $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T(\vec{x}) = \vec{0}_{\mathbb{R}^m}$$

for all $\vec{x} \in \mathbb{R}^n$ is called a **zero transformation**.

Note that there are infinitely many zero transformations, one for each pair of positive integers m and n.

Exercise 83

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a zero transformation.

(a) Show that T is a linear transformation.

(b) Find the standard matrix of T.

We next discuss equality of linear transformations.

Definition 5.3.2 **Equality of Linear Transformations**

Let $T, S : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. If $T(\vec{x}) = S(\vec{x})$ for every $\vec{x} \in \mathbb{R}^n$, then we say T and S are **equal** and we write $T = S$. If for some $\vec{x} \in \mathbb{R}^n$ we have that $T(\vec{x}) \neq S(\vec{x})$, then T and S are not equal and we write $T \neq S$.

It's important to note that we have only defined equality for transformations with the same domain and codomain. If T and S have different domains, for instance, then they are never considered equal.

Example 5.3.3

The linear transformations $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ and $S\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined, respectively, by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

are not equal because their domains are different.

The next theorem states that equality of linear transformations is equivalent to equality of their standard matrices.

Theorem 5.3.4

Let $T, S : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Then $T = S$ if and only if $\big[T\big] = \big[S\big]$.

Proof: We have

$$\begin{aligned} T = S &\iff T(\vec{x}) = S(\vec{x}) \text{ for every } \vec{x} \in \mathbb{R}^n \\ &\iff \big[\,T\,\big]\,\vec{x} = \big[\,S\,\big]\,\vec{x} \text{ for every } \vec{x} \in \mathbb{R}^n \\ &\iff \big[\,T\,\big] = \big[\,S\,\big] \text{ by the Matrix Equality Theorem.} \end{aligned}$$

□

Definition 5.3.5 **Addition, Subtraction and Scalar Multiplication of Linear Transformations**

Let $T, S : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations.

We define the **addition** of T and S to be the function $(T + S) : \mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$(T + S)(\vec{x}) = T(\vec{x}) + S(\vec{x})$$

for every $\vec{x} \in \mathbb{R}^n$.

We define the **subtraction** of S from T to be the function $(T - S) : \mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$(T - S)(\vec{x}) = T(\vec{x}) - S(\vec{x})$$

for every $\vec{x} \in \mathbb{R}^n$.

For $c \in \mathbb{R}$, we define the **scalar multiple** cT of T to be the function $cT : \mathbb{R}^n \to \mathbb{R}^m$ by satisfying

$$(cT)(\vec{x}) = cT(\vec{x})$$

for every $\vec{x} \in \mathbb{R}^n$.

As with matrices in $M_{m\times n}(\mathbb{R})$ and vectors in $\mathbb{R}^n$, for $T, S : \mathbb{R}^n \to \mathbb{R}^m$, we have that $T - S = T + (-1)S$.

Example 5.3.6

Let $T, S : \mathbb{R}^3 \to \mathbb{R}^2$ be linear transformations such that

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 - x_2 + x_3 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_3 \\ x_1 + 2x_2 + 3x_3 \end{bmatrix}.$$

Find expressions for $(T + S)\left(\left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]\right)$ and $(-2T)\left(\left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]\right)$.

Solution: For $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right] \in \mathbb{R}^3$ we have

$$(T + S)(\vec{x}) = T(\vec{x}) + S(\vec{x}) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 - x_2 + x_3 \end{bmatrix} + \begin{bmatrix} x_3 \\ x_1 + 2x_2 + 3x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 + x_3 \\ 2x_1 + x_2 + 4x_3 \end{bmatrix}$$

and

$$(-2)T(\vec{x}) = -2\begin{bmatrix} 2x_1 + x_2 \\ x_1 - x_2 + x_3 \end{bmatrix} = \begin{bmatrix} -4x_1 - 2x_2 \\ -2x_1 + 2x_2 - 2x_3 \end{bmatrix}.$$

It is not difficult to show that the functions $T + S$ and $-2T$ derived in Example 5.3.6 are both linear transformations. Computing the standard matrices for T and S gives

$$\big[\,T\,\big] = \begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \quad \text{and} \quad \big[\,S\,\big] = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}$$

and computing the standard matrices for $T+S$ and $-2T$ shows us that

$$\left[T+S\right]=\begin{bmatrix}2&1&1\\2&1&4\end{bmatrix}=\left[T\right]+\left[S\right]\quad\text{and}\quad\left[-2T\right]=\begin{bmatrix}-4&-2&0\\-2&2&-2\end{bmatrix}=-2\left[T\right].$$

Theorem 5.3.7 Let $T,S:\mathbb{R}^n\to\mathbb{R}^m$ be linear transformations and $c\in\mathbb{R}$. Then

$$T+S:\mathbb{R}^n\to\mathbb{R}^m\quad\text{and}\quad cT:\mathbb{R}^n\to\mathbb{R}^m$$

are linear transformations. Moreover,

$$\left[T+S\right]=\left[T\right]+\left[S\right]\quad\text{and}\quad\left[cT\right]=c\left[T\right].$$

Proof: We prove the result for cT. For any $\vec{x},\vec{y}\in\mathbb{R}^n$ and any $c_1,c_2\in\mathbb{R}$, we have

$$\begin{aligned}(cT)(c_1\vec{x}+c_2\vec{y})&=cT(c_1\vec{x}+c_2\vec{y}) && \text{by definition of } cT\\&=c\big(c_1T(\vec{x})+T(c_2\vec{y})\big) && \text{since } T \text{ is linear}\\&=c_1cT(\vec{x})+c_2cT(\vec{y})\\&=c_1(cT)(\vec{x})+c_2(cT)(\vec{y}) && \text{by definition of } cT\end{aligned}$$

which shows that cT is linear. Now for any $\vec{x}\in\mathbb{R}^n$,

$$\begin{aligned}[cT]\vec{x}&=(cT)(\vec{x})\\&=cT(\vec{x}) && \text{by definition of } cT\\&=c\left[T\right]\vec{x}\end{aligned}$$

from which we see that $\left[cT\right]=c\left[T\right]$ by the Matrix Equality Theorem. $\square$

Exercise 84 Let $T,S:\mathbb{R}^n\to\mathbb{R}^m$ be linear transformations and let $c,d\in\mathbb{R}$. Use Theorem 5.3.7 to show that

(a) $(cT+dS):\mathbb{R}^n\to\mathbb{R}^m$ is a linear transformation.

(b) $\left[cT+dS\right]=c\left[T\right]+d\left[S\right]$.

Generalizing the preceding Exercise, it follows from Theorem 5.3.7 that for linear transformations $T_1,\dots,T_k:\mathbb{R}^n\to\mathbb{R}^m$ and for scalars $c_1,\dots,c_k$, we have

$$(c_1T_1+\cdots+c_kT_k):\mathbb{R}^n\to\mathbb{R}^m$$

is a linear transformation, and that

$$\left[c_1T_1+\cdots+c_kT_k\right]=c_1\left[T_1\right]+\cdots+c_k\left[T_k\right].$$

Thus, the set of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ is closed under the operations of addition and scalar multiplication, and as a result, is closed under linear combinations.

Example 5.3.8

Let $\vec{d} \in \mathbb{R}^2$ be a nonzero vector, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T(\vec{x}) = 2\,\mathrm{proj}_{\vec{d}}\,\vec{x} - \vec{x}$$

for all $\vec{x} \in \mathbb{R}^2$. Recall from Example 5.2.3 that T is a reflection in the line through the origin with direction vector $\vec{d}$.

(a) Show that T is a linear transformation.

(b) Find the standard matrix of T with $\vec{d} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$.

Solution:

(a) Let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $S(\vec{x}) = \mathrm{proj}_{\vec{d}}\,\vec{x}$ for all $\vec{x} \in \mathbb{R}^2$. We see that S is a linear transformation by Theorem 5.2.1(a). Then for every $\vec{x} \in \mathbb{R}^2$,

$$T(\vec{x}) = 2\,\mathrm{proj}_{\vec{d}}\,\vec{x} - \vec{x} = 2S(\vec{x}) - \mathrm{Id}(\vec{x})$$

so $T = 2S - \mathrm{Id}$. Since both S and Id are linear transformations, it follows from Theorem 5.3.7 that T is linear.

(b) Recall from Example 5.2.2 that for $\vec{d} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$,

$$\big[\,S\,\big] = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix},$$

and it follows from Theorem 5.3.7 that

$$\big[\,T\,\big] = \big[\,2S - \mathrm{Id}\,\big] = 2\,\big[\,S\,\big] - \big[\,\mathrm{Id}\,\big] = 2\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

It is useful to compare the solution of Example 5.3.8 to that of Example 5.2.3.

In line with what we have previously observed with vectors in $\mathbb{R}^n$ and matrices in $M_{m\times n}(\mathbb{R})$, the set of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ behaves well under the operations of addition and scalar multiplication. Recalling the Fundamental Properties of Vector Algebra and the Fundamental Properties of Matrix Algebra, the next theorem should feel very familiar.

Theorem 5.3.9

(Fundamental Properties of Linear Transformations)

Let $T, S, L : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations and let $c, d \in \mathbb{R}$. We have

L1. $T + S : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation — closure under addition

L2. $T + S = S + T$ — addition is commutative

L3. $(T + S) + L = T + (S + L)$ — addition is associative

L4. $cT : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation — closure under scalar multiplication

L5. $c(dT) = (cd)T$ — scalar multiplication is associative

L6. $(c+d)T = cT + dT$ distributive law

L7. $c(T+S) = cT + cS$ distributive law

Aside from adding and scaling linear transformations, we can also compose them. We will see that composition of linear transformations is closely tied to matrix multiplication.

Definition 5.3.10

Composition of Linear Transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations. The **composition** $S \circ T : \mathbb{R}^n \to \mathbb{R}^p$ is the function defined by

$$(S \circ T)(\vec{x}) = S(T(\vec{x}))$$

for every $\vec{x} \in \mathbb{R}^n$.

The composition of two functions is illustrated in Figure 5.3.1. It is important to note that in order for $S \circ T$ to be defined, the domain of S must equal the codomain of T.

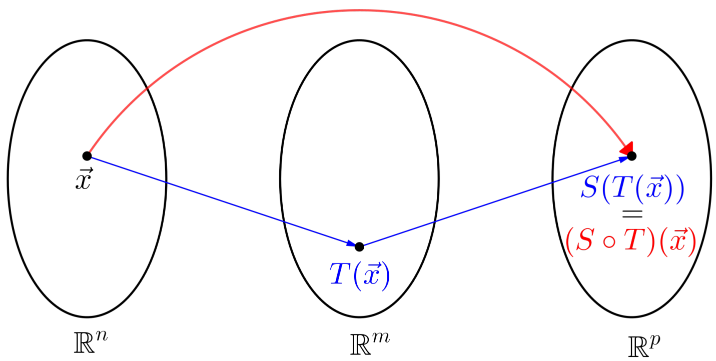


Figure 5.3.1: Composing two functions

Example 5.3.11

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ and $S : \mathbb{R}^2 \to \mathbb{R}^2$ be linear transformations defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ x_2 + x_3 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - 3x_2 \\ 2x_1 \end{bmatrix}.$$

Find an expression for $(S \circ T)\left(\left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]\right)$.

Solution: We have

$$\begin{aligned}(S \circ T)\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= S\left(T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right)\right) \\ &= S\left(\begin{bmatrix} x_1 + x_2 \\ x_2 + x_3 \end{bmatrix}\right)\end{aligned}$$

$$= \begin{bmatrix} (x_1 + x_2) - 3(x_2 + x_3) \\ 2(x_1 + x_2) \end{bmatrix}$$
$$= \begin{bmatrix} x_1 - 2x_2 - 3x_3 \\ 2x_1 + 2x_2 \end{bmatrix}.$$

Exercise 85 Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ and $S : \mathbb{R}^3 \to \mathbb{R}^2$ be linear transformations defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_2 \\ x_1 + x_2 \\ x_1 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 - x_3 \end{bmatrix}$$

(a) Find an expression for $(S \circ T)\left(\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]\right)$.

(b) Find an expression for $(T \circ S)\left(\left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]\right)$.

Notice that in Example 5.3.11, $S \circ T$ is also a linear transformation with domain $\mathbb{R}^3$ and codomain $\mathbb{R}^2$. Its standard matrix is

$$\big[S \circ T\big] = \begin{bmatrix} 1 & -2 & -3 \\ 2 & 2 & 0 \end{bmatrix}.$$

To relate this back to the standard matrices for S and T, which are given by

$$\big[S\big] = \begin{bmatrix} 1 & -3 \\ 2 & 0 \end{bmatrix} \quad \text{and} \quad \big[T\big] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

observe that

$$\big[S\big]\,\big[T\big] = \begin{bmatrix} 1 & -3 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -2 & -3 \\ 2 & 2 & 0 \end{bmatrix} = \big[S \circ T\big]$$

which is the standard matrix for $S \circ T$. That is, we have $[S \circ T] = [S][T]$ – or, in words, the standard matrix of the composition of S and T is the product of the standard matrices of S and T. This is true in general, as next theorem shows.

Theorem 5.3.12 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations. Then

$$S \circ T : \mathbb{R}^n \to \mathbb{R}^p$$

is a linear transformation and

$$\big[S \circ T\big] = \big[S\big]\,\big[T\big].$$

Proof: We first show that $S \circ T$ is linear. Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$. Then

$$\begin{aligned}(S \circ T)(c_1\vec{x} + c_2\vec{y}) &= S\big(T(c_1\vec{x} + c_2\vec{y})\big) \\ &= S\big(c_1T(\vec{x}) + c_2T(\vec{y})\big) && \text{since } T \text{ is linear}\end{aligned}$$

$$
\begin{aligned}
&= c_1 S\big(T(\vec{x})\big) + c_2 S\big(T(\vec{y})\big) && \text{since } S \text{ is linear}\\
&= c_1 (S \circ T)(\vec{x}) + c_2 (S \circ T)(\vec{y})
\end{aligned}
$$

which shows that $S \circ T$ is linear. Now for any $\vec{x} \in \mathbb{R}^n$,

$$
\begin{aligned}
\big[\, S \circ T \,\big]\, \vec{x} &= (S \circ T)(\vec{x})\\
&= S\big(T(\vec{x})\big)\\
&= S\left(\big[\, T \,\big]\, \vec{x}\right)\\
&= \big[\, S \,\big]\left(\big[\, T \,\big]\, \vec{x}\right)\\
&= \left(\big[\, S \,\big]\, \big[\, T \,\big]\right) \vec{x}
\end{aligned}
$$

from which we see that $\big[\, S \circ T \,\big] = \big[\, S \,\big]\, \big[\, T \,\big]$ by the Matrix Equality Theorem. □

Example 5.3.13 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a counterclockwise rotation about the origin by an angle of $\pi/4$ and let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be a projection onto the x_1-axis. Find the standard matrices for $S \circ T$ and $T \circ S$.

Solution: We have

$$
\begin{aligned}
\big[\, T \,\big] &= \begin{bmatrix} \cos\pi/4 & -\sin\pi/4 \\ \sin\pi/4 & \cos\pi/4 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}\\
\big[\, S \,\big] &= \big[\operatorname{proj}_{\vec{e}_1} \vec{e}_1 \quad \operatorname{proj}_{\vec{e}_1} \vec{e}_2\big] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}
\end{aligned}
$$

and thus

$$
\begin{aligned}
\big[\, S \circ T \,\big] = \big[\, S \,\big]\, \big[\, T \,\big] &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ 0 & 0 \end{bmatrix}\\
\big[\, T \circ S \,\big] = \big[\, T \,\big]\, \big[\, S \,\big] &= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & 0 \end{bmatrix}.
\end{aligned}
$$

We notice in the previous example that although $S \circ T$ and $T \circ S$ are both defined, $\big[\, S \circ T \,\big] \neq \big[\, T \circ S \,\big]$ from which we conclude that $S \circ T$ and $T \circ S$ are not the same linear transformation, that is, T and S do not commute under composition. This shouldn't be surprising for two reasons: first, the composition of linear transformations corresponds to multiplication of matrices, and multiplication of matrices is not commutative; and second, you have seen in your calculus courses that composition of functions does not commute. For example, if $f(x) = \sqrt{x}$ and $g(x) = \sin(x)$, then

$$
f(g(x)) = \sqrt{\sin(x)} \neq \sin\left(\sqrt{x}\right) = g(f(x)).
$$

What we've discovered is that, geometrically, performing a rotation followed by a projection will generally not give the same result as performing the same projection followed by the same rotation. Perhaps you can convince yourself that this is true by thinking about it for a bit. However, notice the power of using matrices: this result follows immediately from the straightforward calculation that $[S][T] \neq [T][S]$!

Example 5.3.14

Let $T, S : \mathbb{R}^2 \to \mathbb{R}^2$ be linear transformations defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + x_2 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ -x_1 + 2x_2 \end{bmatrix}.$$

Find $\begin{bmatrix} S \circ T \end{bmatrix}$ and $\begin{bmatrix} T \circ S \end{bmatrix}$.

Solution: Since T and S are linear, we have

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$
$$\begin{bmatrix} S \end{bmatrix} = \begin{bmatrix} S(\vec{e}_1) & S(\vec{e}_2) \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

and thus

$$\begin{bmatrix} S \circ T \end{bmatrix} = \begin{bmatrix} S \end{bmatrix}\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} T \circ S \end{bmatrix} = \begin{bmatrix} T \end{bmatrix}\begin{bmatrix} S \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We see that $\begin{bmatrix} S \circ T \end{bmatrix} = I = \begin{bmatrix} T \circ S \end{bmatrix}$, so $S \circ T = T \circ S$.

Example 5.3.14 shows that $\begin{bmatrix} T \end{bmatrix}$ and $\begin{bmatrix} S \end{bmatrix}$ are inverses of each other. As we will see, this will imply that T and S are inverses of one another.

Exercise 86

With $\vec{d} = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\vec{x}) = \operatorname{proj}_{\vec{d}} \vec{x}$ for all $\vec{x} \in \mathbb{R}^2$. Compute $\begin{bmatrix} T \end{bmatrix}$ and $\begin{bmatrix} T \circ T \end{bmatrix}$, and deduce that $T \circ T = T$.

Compare your solution to the solution of Problem 1.7.6.

The next theorem summarizes the basic properties of compositions of linear transformations.

Theorem 5.3.15

(Properties of Composition)

Let T, S, L be linear transformations with appropriate domains and codomains, and let $c \in \mathbb{R}$. Then:

(a) $\mathrm{Id} \circ T = T$. Id is an identity transformation

(b) $T \circ \mathrm{Id} = T$. Id is an identity transformation

(c) $T \circ (S \circ L) = (T \circ S) \circ L$. Composition is associative

(d) $T \circ (S + L) = T \circ S + T \circ L$. Left distributive law

(e) $(S + L) \circ T = S \circ T + L \circ T$. Right distributive law

(f) $(cT) \circ S = c(T \circ S) = T \circ (cS)$.

In light of Theorem 5.3.12, it is not surprising that Theorem 5.3.15 is so similar to Theorem 3.4.8. This again illustrates the very close connection between matrices and linear transformations. We do not prove Theorem 5.3.15, but you are encouraged to try writing your own proofs.

Section 5.3 Problems

5.3.1. For each of the following linear transformations T and S, compute $(2T+3S)(\vec{x})$.

(a) $T, S : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -2x_1 - x_2 \\ -x_1 - x_2 \end{bmatrix},$$
$$S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -x_1 + x_2 \\ x_1 - 2x_2 \end{bmatrix}.$$

(b) $T, S : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 3x_1 + 2x_2 + 6x_3 \\ 2x_1 + 3x_2 + 5x_3 \\ x_1 + x_2 + 2x_3 \end{bmatrix},$$
$$S\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -x_1 - 2x_2 + 8x_3 \\ -x_1 + 3x_3 \\ x_1 + x_2 - 5x_3 \end{bmatrix}.$$

5.3.2. Let the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a counterclockwise rotation about the origin by $\pi/3$ radians and let the linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ be a reflection in the line $x_2 = x_1$.

(a) Find the standard matrix $\big[T\big]$ of T.

(b) Find the standard matrix $\big[S\big]$ of S.

(c) Find the standard matrix $\big[T \circ S\big]$ of $T \circ S$.

(d) Find the standard matrix $\big[S \circ T\big]$ of $S \circ T$.

5.3.3. Let the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a shear in the x_1-direction by a factor of 5 and let the linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ compression in the x_2-direction by a factor of $\frac{1}{2}$

(a) Find the standard matrix $\big[T\big]$ of T.

(b) Find the standard matrix $\big[S\big]$ of S.

(c) Find the standard matrix of T followed by S.

(d) Find the standard matrix of T following S.

5.4 Inverses of Linear Transformations

Our study of linear transformations has relied heavily on our knowledge of matrix algebra, and as a result, we have gained a geometric intuition of the matrix–vector product and more generally, matrix multiplication. Recall that matrix multiplication led to the notion of an invertible matrix, so it natural that we study invertible linear transformations here. The idea is similar to that of matrices: given two linear transformations T, S, we check whether $S \circ T = \mathrm{Id} = T \circ S$. In order for $S \circ T$ and $T \circ S$ to be equal (with their common value being the identity transformation), we require $\mathbb{R}^n$ to be the domain and codomain of both T and S.

Definition 5.4.1

Invertible Linear Transformation, Inverse Linear Transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. If there exists another linear transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T \circ S = \mathrm{Id} = S \circ T,$$

then T is **invertible** and S is an **inverse** of T (and S is invertible with T an inverse of S).

Our definition refers to S as *an* inverse of T, however, suppose that the linear transformations $S_1, S_2 : \mathbb{R}^n \to \mathbb{R}^n$ are inverses of T. Then $S_1 \circ T = \mathrm{Id}$ and $T \circ S_2 = \mathrm{Id}$. It follows that

$$S_1 = S_1 \circ \mathrm{Id} = S_1 \circ (T \circ S_2) = (S_1 \circ T) \circ S_2 = \mathrm{Id} \circ S_2 = S_2,$$

showing that T has a unique inverse (if it has one at all).

Definition 5.4.2

T^{-1}

If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation, then we denote its inverse by T^{-1}.

Example 5.4.3

Let $T, S : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + 3x_2 \\ x_1 + 2x_2 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 - 3x_2 \\ -x_1 + 2x_2 \end{bmatrix}.$$

Then for any $\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$,

$$\begin{aligned}(T \circ S)\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) &= T\left(S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)\right) = T\left(\begin{bmatrix} 2x_1 - 3x_2 \\ -x_1 + 2x_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 2(2x_1 - 3x_2) + 3(-x_1 + 2x_2) \\ (2x_1 - 3x_2) + 2(-x_1 + 2x_2) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathrm{Id}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)\end{aligned}$$

and

$$\begin{aligned}(S \circ T)\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) &= S\left(T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)\right) = S\left(\begin{bmatrix} 2x_1 + 3x_2 \\ x_1 + 2x_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 2(2x_1 + 3x_2) - 3(x_1 + 2x_2) \\ -(2x_1 + 3x_2) + 2(x_1 + 2x_2) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathrm{Id}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)\end{aligned}$$

So $T \circ S = \mathrm{Id} = S \circ T$ and thus $T^{-1} = S$ (and $S^{-1} = T$).

Example 5.4.3 shows it is quite tedious to verify that two linear transformations are inverses of one another by directly computing the compositions. The next theorem, which is a natural consequence of Theorem 5.3.12, shows that we can resort to using standard matrices.

Theorem 5.4.4 If $T, S : \mathbb{R}^n \to \mathbb{R}^n$ are linear transformations, then S is the inverse of T if and only if $\left[S\right]$ is the inverse of $\left[T\right]$. In particular, T is invertible (as a linear transformation) if and only if $\left[T\right]$ is invertible (as a matrix).

Proof: We have

$$\begin{aligned} S \text{ is the inverse of } T &\iff S \circ T = \mathrm{Id} = T \circ S \\ &\iff \left[S \circ T\right] = \left[\mathrm{Id}\right] = \left[T \circ S\right] \\ &\iff \left[S\right]\left[T\right] = I = \left[T\right]\left[S\right] \\ &\iff \left[S\right] \text{ is the inverse of } \left[T\right]. \end{aligned}$$

This proves the first part. The second part follows from the first, since if T is invertible with inverse S then $[T]$ will be invertible with inverse $[S]$. Conversely, if $[T]$ is invertible, with inverse $B \in M_{n \times n}(\mathbb{R})$, then T will be invertible with inverse the matrix transformation f_B defined by B. Indeed, $\left[T \circ f_B\right] = \left[T\right]\left[f_B\right] = \left[T\right] B = I$ so $T \circ f_B = \mathrm{Id}$ and similarly $f_B \circ T = \mathrm{Id}$. $\square$

It follows from Theorem 5.4.4 that if T is an invertible linear operator on $\mathbb{R}^n$, then

$$\left[T^{-1}\right] = \left[T\right]^{-1}.$$

Example 5.4.5 Let $T, S : \mathbb{R}^2 \to \mathbb{R}^2$ be linear transformations defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + x_2 \end{bmatrix} \quad \text{and} \quad S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ -x_1 + 2x_2 \end{bmatrix}.$$

In Example 5.3.14, we saw that

$$\left[T\right] = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \left[S\right] = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

were inverse matrices, that is, $\left[T\right]^{-1} = \left[S\right]$. By Theorem 5.4.4, we have that $T^{-1} = S$.

Exercise 87 Let T and S be defined as in Example 5.4.3. Use Theorem 5.4.4 to show that $T^{-1} = S$.

Geometrically, given an invertible linear operator T on $\mathbb{R}^n$, we can view $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ as "undoing" what T does.

Example 5.4.6 Recall that $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ denotes a counterclockwise rotation about the origin through an angle of θ. Describe the inverse transformation of R_θ and find its standard matrix.

Solution: The inverse of a counterclockwise rotation by an angle of θ is a counterclockwise rotation by an angle of $-\theta$ (that is, a clockwise rotation by an angle of θ). Thus, the inverse transformation of R_θ is $R_\theta^{-1} = R_{-\theta}$. As we have seen following Example 5.2.6,

$$\left[R_\theta^{-1}\right] = \left[R_{-\theta}\right] = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Note that we have just shown that $\left[R_\theta\right]^{-1} = \left[R_{-\theta}\right]$, that is,

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

We could have used the Matrix Inversion Algorithm to compute $\left[R_\theta\right]^{-1}$, but this would have required us to row reduce

$$\left[\begin{array}{cc|cc} \cos\theta & -\sin\theta & 1 & 0 \\ \sin\theta & \cos\theta & 0 & 1 \end{array}\right] \longrightarrow \left[\begin{array}{cc|cc} 1 & 0 & \cos\theta & \sin\theta \\ 0 & 1 & -\sin\theta & \cos\theta \end{array}\right],$$

which is quite tedious. Indeed, understanding what multiplication by a square matrix does geometrically can give us a fast way to decide if the matrix is invertible, and if so, what the inverse of that matrix is.

Exercise 88

Recall Example 5.2.4. The linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined there is a reflection in the plane with scalar equation $x_1 - x_2 + 2x_3 = 0$ and has standard matrix

$$\left[T\right] = \begin{bmatrix} 2/3 & 1/3 & -2/3 \\ 1/3 & 2/3 & 2/3 \\ -2/3 & 2/3 & -1/3 \end{bmatrix}.$$

Find $\left[T\right]^{-1}$.

Example 5.4.7

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \left(\begin{bmatrix} 2x_1 + 5x_2 \\ x_1 + 3x_2 \end{bmatrix}\right)$.

Find T^{-1}, that is, find an expression for $T^{-1}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$.

Solution: We have

$$\left[T\right] = \left[T(\vec{e}_1) \;\; T(\vec{e}_2)\right] = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}.$$

Applying the Matrix Inversion Algorithm gives

$$\left[\begin{array}{cc|cc} 2 & 5 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{cc|cc} 1 & 3 & 0 & 1 \\ 2 & 5 & 1 & 0 \end{array}\right] \underset{R_2 - 2R_1}{\longrightarrow} \left[\begin{array}{cc|cc} 1 & 3 & 0 & 1 \\ 0 & -1 & 1 & -2 \end{array}\right] \underset{-R_2}{\longrightarrow}$$
$$\left[\begin{array}{cc|cc} 1 & 3 & 0 & 1 \\ 0 & 1 & -1 & 2 \end{array}\right] \xrightarrow{R_1 - 3R_2} \left[\begin{array}{cc|cc} 1 & 0 & 3 & -5 \\ 0 & 1 & -1 & 2 \end{array}\right].$$

Thus

$$\left[T^{-1}\right] = \left[T\right]^{-1} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix},$$

so

$$T^{-1}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 - 5x_2 \\ -x_1 + 2x_2 \end{bmatrix},$$

Exercise 89 The $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_3 \\ x_1 + x_2 \\ x_1 + x_2 + x_3 \end{bmatrix}.$$

Find an expression for $T^{-1}\left(\left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]\right)$.

Section 5.4 Problems

5.4.1. Show that the linear transformations T and S are inverses of one another.

(a) $T, S : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -2x_1 - x_2 \\ -x_1 - x_2 \end{bmatrix},$$
$$S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -x_1 + x_2 \\ x_1 - 2x_2 \end{bmatrix}.$$

(b) $T, S : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 3x_1 + 2x_2 + 6x_3 \\ 2x_1 + 3x_2 + 5x_3 \\ x_1 + x_2 + 2x_3 \end{bmatrix},$$
$$S\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -x_1 - 2x_2 + 8x_3 \\ -x_1 + 3x_3 \\ x_1 + x_2 - 5x_3 \end{bmatrix}.$$

5.4.2. For each linear transformation T, compute $T^{-1}(\vec{x})$.

(a) $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_1 + x_2 \end{bmatrix}$

(b) $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 2x_2 + x_3 \\ x_1 + 5x_2 + 3x_3 \\ -3x_2 - 2x_3 \end{bmatrix}.$

5.4.3. (a) Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Prove that if $T(\vec{x}) = \vec{0}$, then $\vec{x} = \vec{0}$.

(b) Give an example of a non-invertible linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ and a non-zero $\vec{x} \in \mathbb{R}^n$ such that $T(\vec{x}) = \vec{0}$.

5.4.4. Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Let $\vec{y} \in \mathbb{R}^n$ be a fixed vector. Show that there is a vector $\vec{x} \in \mathbb{R}^n$ (that may depend on $\vec{y}$) such that $T(\vec{x}) = \vec{y}$.

5.5 The Kernel and the Range

In mathematics, finding the roots (or zeros) of a function f, that is, solving $f(x) = 0$, is a very common and necessary practice. In calculus, for example, the roots of the derivative f' of f are important when determining the local minima and maxima of f. Unfortunately, finding roots of a function can become extremely difficult, if not impossible, when the expression for the function becomes complicated. As we will see in this section, determining the roots of linear transformations is quite straightforward.

Definition 5.5.1
Kernel of a Linear Transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **kernel** of T is

$$\mathrm{Ker}(T) = \{\vec{x} \in \mathbb{R}^n \,|\, T(\vec{x}) = \vec{0}\}.$$

Note that $\mathrm{Ker}(T) \subseteq \mathbb{R}^n$, that is, $\mathrm{Ker}(T)$ is a subset of the domain of T. The kernel of T is also sometimes called the *nullspace* of T, denoted by $\mathrm{Null}(T)$.

Example 5.5.2

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ -3x_1 + 3x_2 \end{bmatrix}.$$

Determine which of $\vec{x}_1 = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, $\vec{x}_2 = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $\vec{x}_3 = \left[\begin{smallmatrix} 3 \\ 2 \end{smallmatrix}\right]$ belong to $\mathrm{Ker}(T)$.

Solution: We compute

$$\begin{aligned} T(\vec{x}_1) &= T\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 - 0 \\ -3(0) + 3(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ T(\vec{x}_2) &= T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 - 1 \\ -3(1) + 3(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ T(\vec{x}_3) &= T\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 3 - 2 \\ -3(3) + 3(2) \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \end{aligned}$$

from which we deduce that $\vec{x}_1, \vec{x}_2 \in \mathrm{Ker}(T)$ and $\vec{x}_3 \notin \mathrm{Ker}(T)$.

Exercise 90

Consider the following linear transformations:

$$\begin{aligned} T_1\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= \begin{bmatrix} 2x_1 - x_2 \\ 3x_2 - 2x_3 \\ x_1 + x_2 - x_3 \end{bmatrix} \\ T_2\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= \begin{bmatrix} x_1 - 5x_2 + 4x_3 \\ 0 \end{bmatrix} \\ T_3\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= 5x_1 - 4x_2 + x_3. \end{aligned}$$

Determine which of $\mathrm{Ker}(T_1)$, $\mathrm{Ker}(T_2)$ and $\mathrm{Ker}(T_3)$ contain $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Given a function f, one is also concerned with the collection of "outputs" of that function, that is, the set of all possible values of $f(x)$. For example, if v is a function that models the speed of a car at any given time t, then we may be interested in determining at which times the car reaches a given speed, or we may wish to know what possible speeds the car attains during a given period of time. Answering such questions requires knowledge of the *range* of the function. This section will also address how to find the range of a linear transformation, and as with the kernel, we will see that this is a straightforward process.

Definition 5.5.3

Range of a Linear Transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **range** of T is

$$\mathrm{Range}(T) = \{T(\vec{x}) \mid \vec{x} \in \mathbb{R}^n\}.$$

Note that $\mathrm{Range}(T) \subseteq \mathbb{R}^m$, that is, $\mathrm{Range}(T)$ is a subset of the codomain of T. Figure 5.5.1 gives a helpful visualization of the kernel and range of a linear transformation.

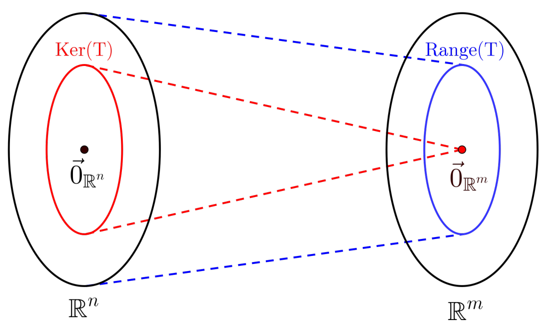


Figure 5.5.1: Visualizing the kernel and the range of a linear transformation with domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$.

Example 5.5.4

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ 2x_1 + x_2 \\ 3x_2 \end{bmatrix}.$$

Determine which of $\vec{y}_1 = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$ and $\vec{y}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ belong to $\mathrm{Range}(T)$.

Solution: To see if $\vec{y}_1 \in \mathrm{Range}(T)$, we try to find $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$ such that $T(\vec{x}) = \vec{y}_1$. Thus we need

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ 2x_1 + x_2 \\ 3x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}.$$

This leads to a system of equations

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & 2 \\ 2x_1 & + & x_2 & = & 3 \\ & & 3x_2 & = & 3 \end{array}$$

Carrying the augmented matrix of this system to reduced row echelon form gives

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 0 & 3 & 3 \end{array}\right] \underset{R_2-2R_1}{\longrightarrow} \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & -1 & -1 \\ 0 & 3 & 3 \end{array}\right] \underset{R_3+3R_1}{\overset{R_1+R_2}{\longrightarrow}} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{array}\right] \underset{-R_2}{\longrightarrow} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right]$$

from which we see that $x_1 = x_2 = 1$ and so $T(1,1) = (2,3,3)$. Thus $\vec{y}_1 \in \text{Range}(T)$.

For $\vec{y}_2$, we seek $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$ such that

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

A similar computation leads to a system of equations with augmented matrix

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 0 & 3 & 2 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{array}\right].$$

As this system is inconsistent, there is no $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$ such that $T(\vec{x}) = \vec{y}_2$ and so $\vec{y}_2 \notin \text{Range}(T)$.

Exercise 91 Consider the following linear transformations:

$$T_1\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 - x_3 \\ x_2 - 2x_3 \\ -2x_1 - x_3 \end{bmatrix} \quad \text{and} \quad T_2\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 - 3x_2 \\ x_1 + x_2 \\ 2x_1 - x_2 \end{bmatrix}.$$

Determine which of $\text{Range}(T_1)$ and $\text{Range}(T_2)$ contain $\vec{y} = \left[\begin{smallmatrix} 1 \\ 3 \\ 3 \end{smallmatrix}\right]$.

Note that in Example 5.5.4, to see if a vector $\vec{y} \in \text{Range}(T)$, were are ultimately checking if the linear system of equations $\big[T\big]\,\vec{x} = \vec{y}$ is consistent, that is, if $\vec{y} \in \text{Col}\left(\big[T\big]\right)$. Recalling that for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(\vec{x}) = \big[T\big]\,\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$, the following theorem should not be too surprising. However, it is a very important theorem since it allows us to reduce the problems of determining $\text{Ker}(T)$ and $\text{Range}(T)$ to problems involving the matrix $[T]$ that we have already learned how to solve.

Theorem 5.5.5 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix $\big[T\big]$. Then

(a) $\text{Ker}(T) = \text{Null}\left(\big[T\big]\right)$, and

(b) $\text{Range}(T) = \text{Col}\left(\left[\,T\,\right]\right)$.

In particular, $\text{Ker}(T)$ is a subspace of $\mathbb{R}^n$ and $\text{Range}(T)$ is a subspace of $\mathbb{R}^m$.

Proof:

(a) Since

$$\vec{x} \in \text{Ker}(T) \iff T(\vec{x}) = \vec{0} \iff \left[\,T\,\right]\vec{x} = \vec{0} \iff \vec{x} \in \text{Null}\left(\left[\,T\,\right]\right),$$

we have that $\text{Ker}(T) = \text{Null}\left(\left[\,T\,\right]\right)$ and thus $\text{Ker}(T)$ is a subspace of $\mathbb{R}^n$.

(b) Since

$$\begin{aligned}\vec{y} \in \text{Range}(T) &\iff \vec{y} = T(\vec{x}) \text{ for some } \vec{x} \in \mathbb{R}^n \\ &\iff \vec{y} = \left[\,T\,\right]\vec{x} \text{ for some } \vec{x} \in \mathbb{R}^n \\ &\iff \vec{y} \in \text{Col}\left(\left[\,T\,\right]\right),\end{aligned}$$

we see that $\text{Range}(T) = \text{Col}\left(\left[\,T\,\right]\right)$ and thus $\text{Range}(T)$ is a subspace of $\mathbb{R}^m$. □

Exercise 92 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Without referring to Theorem 5.5.5, prove that

(a) $\text{Ker}(T)$ is a subspace of $\mathbb{R}^n$, and

(b) $\text{Range}(T)$ is a subspace of $\mathbb{R}^m$.

Using Theorem 5.5.5, we now have a method for determining $\text{Ker}(T)$ and $\text{Range}(T)$ for any linear transformation T: first find the standard matrix $\left[\,T\,\right]$ of T, and then compute $\text{Null}\left(\left[\,T\,\right]\right)$ and $\text{Col}\left(\left[\,T\,\right]\right)$. We've already talked about how to find nullspaces and column spaces of matrices in Section 4.6, so it might be a good idea to review that section now to refresh your memory.

Example 5.5.6 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a projection onto the line through the origin with direction vector $\vec{d} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Find a basis for $\text{Ker}(T)$ and $\text{Range}(T)$.

Solution: The standard matrix of T is

$$\left[\,T\,\right] = \left[T(\vec{e}_1)\ \ T(\vec{e}_2)\ \ T(\vec{e}_3)\right] = \left[\text{proj}_{\vec{d}}\ \vec{e}_1\ \ \text{proj}_{\vec{d}}\ \vec{e}_2\ \ \text{proj}_{\vec{d}}\ \vec{e}_3\right] = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

To find a basis for $\text{Ker}(T)$, we solve the homogeneous system of equations given by $\left[\,T\,\right]\vec{x} = \vec{0}$. Carrying $\left[\,T\,\right]$ to reduced row echelon form gives

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{matrix} \longrightarrow \\ R_2 - R_1 \\ R_3 - R_1 \end{matrix} \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} 3R_1 \\ \longrightarrow \\ \end{matrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and we see that

$$\vec{x} = s\begin{bmatrix}-1\\1\\0\end{bmatrix} + t\begin{bmatrix}-1\\0\\1\end{bmatrix}, \quad s,t\in\mathbb{R}$$

so

$$\left\{\begin{bmatrix}-1\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}\right\}$$

is a basis for $\mathrm{Ker}(T)$. From our work above, we see that the reduced row echelon form of $\big[\,T\,\big]$ has a leading one in the first column only, and so the Extraction Theorem gives that a basis for $\mathrm{Range}(T)$ is

$$\left\{\begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix}\right\}.$$

In Example 5.5.6, note that geometrically, $\mathrm{Ker}(T)$ is a plane through the origin (two-dimensional subspace) in $\mathbb{R}^3$, and that $\mathrm{Range}(T)$ is a line through the origin (one-dimensional subspace) in $\mathbb{R}^3$ with direction vector $\vec{d} = \left[\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right]$.

Exercise 93 Find a basis for $\mathrm{Ker}(T)$ and $\mathrm{Range}(T)$ where T is the linear transformation defined by

$$T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}x_1+x_2\\x_1+x_2+x_3\end{bmatrix}.$$

We end this chapter by revisiting the Rank–Nullity Theorem. For a linear transformation $T:\mathbb{R}^n\to\mathbb{R}^m$, we know that $\big[\,T\,\big]\in M_{m\times n}(\mathbb{R})$. It follows from the Rank–Nullity Theorem that

$$\mathrm{rank}\left(\big[\,T\,\big]\right) + \mathrm{nullity}\left(\big[\,T\,\big]\right) = n,$$

that is,

$$\dim\left(\mathrm{Col}\left(\big[\,T\,\big]\right)\right) + \dim\left(\mathrm{Null}\left(\big[\,T\,\big]\right)\right) = n.$$

In light of Theorem 5.5.5, we have the following result.

Theorem 5.5.7 For any linear transformation $T:\mathbb{R}^n\to\mathbb{R}^m$, we have

$$\dim(\mathrm{Range}(T)) + \dim(\mathrm{Ker}(T)) = n.$$

Example 5.5.8 Let $T:\mathbb{R}^6\to\mathbb{R}^9$ be a linear transformation such that

$$B = \left\{\begin{bmatrix}1\\0\\1\\2\\1\\0\end{bmatrix}, \begin{bmatrix}1\\2\\1\\5\\3\\4\end{bmatrix}\right\}$$

is a basis for $\mathrm{Ker}(T)$. Determine $\dim(\mathrm{Range}(T))$.

Solution: Since B contains two vectors, we see that $\dim(\mathrm{Ker}(T)) = 2$. It then follows from Theorem 5.5.7 that $2 + \dim(\mathrm{Range}(T)) = 6$, so $\dim(\mathrm{Range}(T)) = 4$.

Exercise 94 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that if $\mathrm{Range}(T) = \mathbb{R}^m$, then $m \leq n$.

Section 5.5 Problems

5.5.1. Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the linear transformation given by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + 2x_2 \\ x_1 + x_3 \\ x_2 + x_4 \end{bmatrix}.$$

Find a basis for $\mathrm{Ker}(T)$ and $\mathrm{Range}(T)$, and state their dimensions.

5.5.2. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_2 + x_3 \\ -x_1 + 2x_3 \\ -x_1 - 2x_2 \end{bmatrix}.$$

Find a basis for $\mathrm{Ker}(T)$ and $\mathrm{Range}(T)$, and state their dimensions.

5.5.3. Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Prove that $\mathrm{Ker}(T) = \{\vec{0}\}$ if and only if $\mathrm{Range}(T) = \mathbb{R}^n$.

Chapter 6

Determinants

In this chapter, we discuss a number, called the determinant, that is associated to a real square matrix, that is, to a matrix $A \in M_{n\times n}(\mathbb{R})$. We will examine how to compute determinants and will see that a matrix is invertible if and only if its determinant is nonzero. We will also examine how the determinant can be used to determine areas of parallelograms and volumes of parallelepipeds.

6.1 Determinants and Invertibility

Let $A \in M_{n\times n}(\mathbb{R})$. The invertibility of A was discussed in Section 3.5. There, the Matrix Inversion Algorithm was introduced, which allows us to both determine if A is invertible and compute A^{-1} if A is in fact invertible. This section will examine another way to determine if a matrix A is invertible.

Example 6.1.1 Let $A = \begin{bmatrix} a \end{bmatrix} \in M_{1\times 1}(\mathbb{R})$. Then by the Matrix Invertibility Criteria, A is invertible if and only if $\operatorname{rank}(A) = 1$. But clearly, $\operatorname{rank}(A) = 1$ if and only if $a \neq 0$. Thus A is invertible if and only if $a \neq 0$.

Example 6.1.2 Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$. By the Matrix Invertibility Criteria, A is invertible if and only if $\operatorname{rank}(A) = 2$. In order for $\operatorname{rank}(A) = 2$, we require that at least one of a and c be nonzero. Assume that $a \neq 0$. Then carrying A to row echelon form gives

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \underset{R_2 - \frac{c}{a}R_1}{\longrightarrow} \begin{bmatrix} a & b \\ 0 & d - \frac{bc}{a} \end{bmatrix}.$$

For $\operatorname{rank}(A) = 2$, we require $d - \frac{bc}{a} \neq 0$, that is, we require $ad - bc \neq 0$. Hence A is invertible if and only if $ad - bc \neq 0$. Note that we would have arrived at the same conclusion had we instead assumed $c \neq 0$.

Example 6.1.1 and Example 6.1.2 show that we can look at the entries of a 1×1 or 2×2 matrix to determine if that matrix is invertible. This leads us to make the following definition.

Definition 6.1.3

1×1 Determinant, 2×2 Determinant

For $A = \begin{bmatrix} a \end{bmatrix} \in M_{1\times 1}(\mathbb{R})$, the **determinant** of A is

$$\det(A) = \det\left(\begin{bmatrix} a \end{bmatrix}\right) = a,$$

and for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2}(\mathbb{R})$, the **determinant** of A is

$$\det(A) = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc.$$

For $A \in M_{1\times 1}(\mathbb{R})$ or for $A \in M_{2\times 2}(\mathbb{R})$, it now follows from Example 6.1.1 and Example 6.1.2 that A is invertible if and only if $\det(A) \neq 0$.

Example 6.1.4

Let $A = \begin{bmatrix} -3 \end{bmatrix}$. Then $\det(A) = \det\left(\begin{bmatrix} -3 \end{bmatrix}\right) = -3$. Since $\det(A) \neq 0$, A is invertible.

Example 6.1.5

Consider $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Then

$$\det(A) = \det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) = 1(4) - 2(3) = 4 - 6 = -2.$$

Since $\det(A) \neq 0$, A is invertible.

Exercise 95

Let $A = \begin{bmatrix} 2 & -1 \\ 5 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -6 \\ -1 & 2 \end{bmatrix}$. Compute $\det(A)$ and $\det(B)$ and determine which of A and B are invertible.

It is natural to now extend the definition of a determinant to $n \times n$ matrices. For $n = 3$, consider

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \in M_{3\times 3}(\mathbb{R}).$$

A similar derivation to what was done in Example 6.1.1 and Example 6.1.2 (the details of which we omit) leads us to conclude that A is invertible if and only if

$$aei - afh - bdi + bfg + cdh - ceg \neq 0.$$

We thus define the determinant of A as

$$\det(A) = \det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = aei - afh - bdi + bfg + cdh - ceg.$$

However, we do not make this a formal definition as it is quite difficult to remember and thus not a practical formula to use. In fact, as n increases, defining the determinant of

$A \in M_{n\times n}(\mathbb{R})$ in this way becomes even more cumbersome to write out and impossible to remember.

We instead make the following definition, which will allow us to define the determinant of $A \in M_{n\times n}(\mathbb{R})$ in a more meaningful way.

Definition 6.1.6 **Cofactors**

Let $A \in M_{n\times n}(\mathbb{R})$ with $n \geq 2$ and let $A(i,j)$ be the $(n-1)\times(n-1)$ matrix obtained from A by deleting the ith row and jth column of A. The (i,j)-**cofactor** of A, denoted by $C_{ij}(A)$, is

$$C_{ij}(A) = (-1)^{i+j}\det(A(i,j)),$$

where $i = 1,\dots,n$ and $j = 1,\dots,n$.

Example 6.1.7

Let $A = \begin{bmatrix} 2 & 3 \\ -1 & -5 \end{bmatrix}$. The $(1,1)$-cofactor of A is

$$C_{11}(A) = (-1)^{1+1}\det(A(1,1)) = (-1)^2\det\left(\begin{bmatrix} -5 \end{bmatrix}\right) = -5,$$

and the $(1,2)$-cofactor of A is

$$C_{12}(A) = (-1)^{1+2}\det(A(1,2)) = (-1)^3\det\left(\begin{bmatrix} -1 \end{bmatrix}\right) = 1.$$

Example 6.1.8

Let $A = \begin{bmatrix} 1 & -2 & 3 \\ 1 & 0 & 4 \\ 4 & 1 & 1 \end{bmatrix}$. Then the $(3,2)$-cofactor of A is

$$C_{32}(A) = (-1)^{3+2}\det(A(3,2)) = (-1)^5\det\left(\begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}\right) = -1(4-3) = -1,$$

and the $(2,2)$-cofactor of A is

$$C_{22}(A) = (-1)^{2+2}\det(A(2,2) = (-1)^4\det\left(\begin{bmatrix} 1 & 3 \\ 4 & 1 \end{bmatrix}\right) = 1(1-12) = -11.$$

Exercise 96

Let $A = \begin{bmatrix} 99 & 1 & -1 \\ -100 & 1 & 2 \\ 101 & 2 & 1 \end{bmatrix}$. Determine $C_{11}(A)$, $C_{21}(A)$, and $C_{31}(A)$.

The next example shows how cofactors can be used to compute the determinant of a 2×2 matrix.

Example 6.1.9

Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Then $\det(A) = a_{11}a_{22} - a_{12}a_{21}$. We compute the four cofactors of A:

$$C_{11}(A) = (-1)^{1+1}\det(A(1,1)) = (-1)^2\det\left(\begin{bmatrix} a_{22} \end{bmatrix}\right) = a_{22},$$

$$C_{12}(A) = (-1)^{1+2}\det(A(1,2)) = (-1)^3\det\left(\begin{bmatrix} a_{21} \end{bmatrix}\right) = -a_{21},$$
$$C_{21}(A) = (-1)^{2+1}\det(A(2,1)) = (-1)^3\det\left(\begin{bmatrix} a_{12} \end{bmatrix}\right) = -a_{12},$$
$$C_{22}(A) = (-1)^{2+2}\det(A(2,2)) = (-1)^4\det\left(\begin{bmatrix} a_{11} \end{bmatrix}\right) = a_{11}.$$

Multiplying the entries in the first row of A by the corresponding cofactors and then adding the results gives

$$a_{11}C_{11}(A) + a_{12}C_{12}(A) = a_{11}a_{22} + a_{12}(-a_{21}) = \det(A),$$

and multiplying the entries in the second row of A by the corresponding cofactors and then adding the results gives

$$a_{21}C_{21}(A) + a_{22}C_{22}(A) = a_{21}(-a_{12}) + a_{22}a_{11} = \det(A).$$

Similarly, multiplying the entries in the first column of A by the corresponding cofactors and then adding the results gives

$$a_{11}C_{11}(A) + a_{21}C_{21}(A) = a_{11}a_{22} + a_{21}(-a_{12}) = \det(A),$$

and multiplying the entries in the second column of A by the corresponding cofactors and then adding the results gives

$$a_{12}C_{12}(A) + a_{22}C_{22}(A) = a_{12}(-a_{21}) + a_{22}a_{11} = \det(A).$$

Example 6.1.9 shows that to compute the determinant of a 2×2 matrix, we may pick *any* row (or column) of that matrix, multiply the entries of that row (or column) by the corresponding cofactors and add the results. This motivates the following definition.

Definition 6.1.10

$n \times n$ Determinant, Cofactor Expansion

Let $A = [a_{ij}] \in M_{n\times n}(\mathbb{R})$ with $n \geq 2$. For any $i = 1, \ldots, n$, we define the **determinant** of A as

$$\det(A) = a_{i1}C_{i1}(A) + a_{i2}C_{i2}(A) + \cdots + a_{in}C_{in}(A),$$

which we refer to as a **cofactor expansion along the ith row of** A. Equivalently, for any $j = 1, \ldots, n$,

$$\det(A) = a_{1j}C_{1j}(A) + a_{2j}C_{2j}(A) + \cdots + a_{nj}C_{nj}(A),$$

which we refer to as a **cofactor expansion along the jth column of** A.

It does not matter which row or column is chosen when using a cofactor expansion to compute a determinant of an $n \times n$ matrix. This was verified for the case $n = 2$ in Example 6.1.9, and we omit the verification for the case $n \geq 3$ as it is quite cumbersome.

Having now defined the determinant for any $A \in M_{n\times n}(\mathbb{R})$, we state the main result of this section.

Theorem 6.1.11

Let $A \in M_{n\times n}(\mathbb{R})$. Then A is invertible if and only if $\det(A) \neq 0$.

Theorem 6.1.11 was proven for the cases $n=1$ and $n=2$ in Example 6.1.1 and Example 6.1.2 respectively. We omit the general proof as it is again quite tedious and unenlightening.

Example 6.1.12 Compute $\det(A)$ where $A=\begin{bmatrix}1 & 2 & -3\\ 4 & -5 & 6\\ -7 & 8 & 9\end{bmatrix}$ and determine if A is invertible.

Solution: Performing a cofactor expansion along the first row of A gives

$$\begin{aligned}
\det(A) = \det\left(\begin{bmatrix}1 & 2 & -3\\ 4 & -5 & 6\\ -7 & 8 & 9\end{bmatrix}\right)\\
&= 1C_{11}(A)+2C_{12}(A)-3C_{13}(A)\\
&= 1(-1)^{1+1}\det(A(1,1))+2(-1)^{1+2}\det(A(1,2))-3(-1)^{1+3}\det(A(1,3))\\
&= 1(-1)^2\det\left(\begin{bmatrix}-5 & 6\\ 8 & 9\end{bmatrix}\right)+2(-1)^3\det\left(\begin{bmatrix}4 & 6\\ -7 & 9\end{bmatrix}\right)-3(-1)^4\det\left(\begin{bmatrix}4 & -5\\ -7 & 8\end{bmatrix}\right)\\
&= 1(-45-48)-2(36+42)-3(32-35)\\
&= 1(-93)-2(78)-3(3)\\
&= -93-156+9\\
&= -240.
\end{aligned}$$

Alternatively, a cofactor expansion along the second columns leads to

$$\begin{aligned}
\det(A) = \det\left(\begin{bmatrix}1 & 2 & -3\\ 4 & -5 & 6\\ -7 & 8 & 9\end{bmatrix}\right)\\
&= 2C_{12}(A)-5C_{22}(A)+8C_{32}(A)\\
&= 2(-1)^{1+2}\det(A(1,2))-5(-1)^{2+2}\det(A(2,2))+8(-1)^{3+2}\det(A(3,2))\\
&= 2(-1)^3\det\left(\begin{bmatrix}4 & 6\\ -7 & 9\end{bmatrix}\right)-5(-1)^4\det\left(\begin{bmatrix}1 & -3\\ -7 & 9\end{bmatrix}\right)+8(-1)^5\det\left(\begin{bmatrix}1 & -3\\ 4 & 6\end{bmatrix}\right)\\
&= -2(36+42)-5(9-21)-8(6+12)\\
&= -2(78)-5(-12)-8(18)\\
&= -156+60-144\\
&= -240.
\end{aligned}$$

We see that $\det(A)\neq 0$, so A is invertible.

Exercise 97 Let $A=\begin{bmatrix}a & b & c\\ d & e & f\\ g & h & i\end{bmatrix}$. Show that $\det(A)=aei-afh-bdi+bfg+cdh-ceg$.

We introduce here a convenient notation for the determinant. For a matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \in M_{n\times n}(\mathbb{R}),$$

with $n \geq 2$, we may denote $\det(A)$ by

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}.$$

Thus from Example 6.1.7, we can write

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

and from Example 6.1.12, we have

$$\begin{vmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \\ -7 & 8 & 9 \end{vmatrix} = -240.$$

It is important to *avoid* this notation when computing the determinant of a 1×1 matrix as it can lead to confusion. For example, for $A = \begin{bmatrix} -3 \end{bmatrix}$, we have that $\det A = -3$. If we use the above convention, we would write $\begin{vmatrix} -3 \end{vmatrix} = -3$ which looks like we are saying that the absolute value of -3 is -3. We see that our new notation is ambiguous for 1×1 matrices, and so we do not use it in this case.

The work presented in Example 6.1.12 to evaluate the determinant using a cofactor expansion requires a lot of writing. We present a slightly faster way to write out such solutions. We note that the cofactor $C_{ij}(A)$ is composed of two parts: $(-1)^{i+j}$ and $\det(A(i,j))$. We can write down $A(i,j)$ simply by looking at A and removing the ith row and the jth column. We also realize that $(-1)^{i+j}$ will be either 1 or -1 depending on whether $i+j$ is even or odd. For an $n \times n$ matrix, we can determine the sign of $(-1)^{i+j}$ by simply looking at an $n \times n$ table consisting of "+" and "−" symbols:

$$\begin{matrix} + & - \\ - & + \end{matrix} \quad , \quad \begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix} \quad , \quad \begin{matrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{matrix} \quad , \quad \cdots$$

Notice that we always have a "+" in upper-left corner of the table and we change sign as we move left/right or up/down. To compute $(-1)^{i+j}$, we can simply look to the (i,j)-entry of the appropriately-sized table.

For example, to compute the determinant of $A = \left[\begin{smallmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \\ -7 & 8 & 9 \end{smallmatrix}\right]$ from Example 6.1.12 using a cofactor expansion along the first row, we have

$$\det(A) = \begin{vmatrix} \mathbf{1} & \mathbf{2} & \mathbf{-3} \\ 4 & -5 & 6 \\ -7 & 8 & 9 \end{vmatrix} = \overbrace{+1}^{\left.\begin{smallmatrix} \mathbf{+} & - & + \\ - & + & - \\ + & - & + \end{smallmatrix}\right.} \underbrace{\begin{vmatrix} -5 & 6 \\ 8 & 9 \end{vmatrix}}_{\left|\begin{smallmatrix} \mathbf{1} & 2 & -3 \\ 4 & \mathbf{-5} & \mathbf{6} \\ -7 & \mathbf{8} & \mathbf{9} \end{smallmatrix}\right|} \overbrace{-2}^{\left.\begin{smallmatrix} + & \mathbf{-} & + \\ - & + & - \\ + & - & + \end{smallmatrix}\right.} \underbrace{\begin{vmatrix} 4 & 6 \\ -7 & 9 \end{vmatrix}}_{\left|\begin{smallmatrix} 1 & \mathbf{2} & -3 \\ \mathbf{4} & -5 & \mathbf{6} \\ \mathbf{-7} & 8 & \mathbf{9} \end{smallmatrix}\right|} \overbrace{+(-3)}^{\left.\begin{smallmatrix} + & - & \mathbf{+} \\ - & + & - \\ + & - & + \end{smallmatrix}\right.} \underbrace{\begin{vmatrix} 4 & -5 \\ -7 & 8 \end{vmatrix}}_{\left|\begin{smallmatrix} 1 & 2 & \mathbf{-3} \\ \mathbf{4} & \mathbf{-5} & 6 \\ \mathbf{-7} & \mathbf{8} & 9 \end{smallmatrix}\right|},$$

while using a cofactor expansion along the second column would give

$$\det(A) = \begin{vmatrix} 1 & \mathbf{2} & -3 \\ 4 & \mathbf{-5} & 6 \\ -7 & \mathbf{8} & 9 \end{vmatrix} = \overbrace{-}^{\substack{+\,-\,+\\-\,+\,-\\+\,-\,+}}2 \underbrace{\begin{vmatrix} 4 & 6 \\ -7 & 9 \end{vmatrix}}_{\begin{vmatrix} 1 & \mathbf{2} & -3 \\ \mathbf{4} & -5 & \mathbf{6} \\ \mathbf{-7} & 8 & \mathbf{9} \end{vmatrix}} \overbrace{+}^{\substack{+\,-\,+\\-\,+\,-\\+\,-\,+}}(-5) \underbrace{\begin{vmatrix} 1 & -3 \\ -7 & 9 \end{vmatrix}}_{\begin{vmatrix} \mathbf{1} & 2 & \mathbf{-3} \\ 4 & \mathbf{-5} & 6 \\ \mathbf{-7} & 8 & \mathbf{9} \end{vmatrix}} \overbrace{-}^{\substack{+\,-\,+\\-\,+\,-\\+\,-\,+}}8 \underbrace{\begin{vmatrix} 1 & -3 \\ 4 & 6 \end{vmatrix}}_{\begin{vmatrix} \mathbf{1} & 2 & \mathbf{-3} \\ \mathbf{4} & -5 & \mathbf{6} \\ -7 & \mathbf{8} & 9 \end{vmatrix}}.$$

Hence a more concise solution to Example 6.1.12 using a cofactor expansion along the first row of A is

$$\begin{aligned} \det(A) = \begin{vmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \\ -7 & 8 & 9 \end{vmatrix} &= 1 \begin{vmatrix} -5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ -7 & 9 \end{vmatrix} - 3 \begin{vmatrix} 4 & -5 \\ -7 & 8 \end{vmatrix} \\ &= 1(-45-48) - 2(36+42) - 3(32-35) \\ &= 1(-93) - 2(78) - 3(-3) \\ &= -240. \end{aligned}$$

Exercise 98

Find $\det(B)$ where $B = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 3 & 4 \\ 3 & 6 & 2 \end{bmatrix}$. Is B invertible?

The next example shows that the cofactor expansion quickly becomes inefficient for $n \times n$ matrices when n becomes large.

Example 6.1.13

Let $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$. Evaluate $\det(A)$.

Solution: Performing a cofactor expansion along the first row of A gives

$$\det(A) = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \end{vmatrix} = 1 \begin{vmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \end{vmatrix} + 1 \begin{vmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 1 & 2 & 4 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 3 \end{vmatrix}$$

and then performing cofactor expansions along the first row in each of the resulting determinants gives

$$\begin{aligned} \det(A) &= \left(1 \begin{vmatrix} 2 & 3 \\ 3 & 4 \end{vmatrix} - 1 \begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} + 2 \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix}\right) - \left(1 \begin{vmatrix} 2 & 3 \\ 3 & 4 \end{vmatrix} - 1 \begin{vmatrix} 1 & 3 \\ 1 & 4 \end{vmatrix} + 2 \begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix}\right) \\ &\quad + \left(1 \begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} - 1 \begin{vmatrix} 1 & 3 \\ 1 & 4 \end{vmatrix} + 2 \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix}\right) - \left(1 \begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} - 1 \begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} + 1 \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix}\right) \\ &= \big(-1-(-2)+2(-1)\big) - \big(-1-1+2(1)\big) + \big(-2-1+2(1)\big) - \big(-1-1+1\big) \\ &= -1-0-1-(-1) \\ &= -1. \end{aligned}$$

Example 6.1.13 clearly shows the recursiveness of the cofactor expansion. To compute the determinant of an $n \times n$ matrix, a cofactor expansion (along any row or column) leads to us computing the determinants of n matrices of size $(n-1) \times (n-1)$, and each of these n determinants would require a cofactor expansion as well, which would lead to determinants of $(n-2) \times (n-2)$ matrices and so on. Even on a computer, the cofactor expansion becomes expensive as n becomes large.

As the next example will show, performing a cofactor expansion along a row or column of a matrix that contains many zero entries can greatly reduce the work in computing a determinant.

Example 6.1.14

Determine if $A = \begin{bmatrix} 1 & 2 & -1 & 3 \\ 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 3 \\ -1 & 1 & 2 & 1 \end{bmatrix}$ is invertible by computing $\det(A)$.

Solution: Performing a cofactor expansion along the third row, we have

$$\begin{aligned} \det(A) = \begin{vmatrix} 1 & 2 & -1 & 3 \\ 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 3 \\ -1 & 1 & 2 & 1 \end{vmatrix} &= 0 \begin{vmatrix} 2 & -1 & 3 \\ 2 & 0 & 4 \\ 1 & 2 & 1 \end{vmatrix} - 0 \begin{vmatrix} 1 & -1 & 3 \\ 1 & 0 & 4 \\ -1 & 2 & 1 \end{vmatrix} + 0 \begin{vmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ -1 & 1 & 1 \end{vmatrix} - 3 \begin{vmatrix} 1 & 2 & -1 \\ 1 & 2 & 0 \\ -1 & 1 & 2 \end{vmatrix} \\ &= -3 \begin{vmatrix} 1 & 2 & -1 \\ 1 & 2 & 0 \\ -1 & 1 & 2 \end{vmatrix}. \end{aligned}$$

To evaluate the determinant of the resulting 3×3 matrix, we perform a cofactor expansion along the third column. This gives

$$\begin{aligned} \det(A) &= -3 \left(-1 \begin{vmatrix} 1 & 2 \\ -1 & 1 \end{vmatrix} + 2 \begin{vmatrix} 1 & 2 \\ 1 & 2 \end{vmatrix} \right) \\ &= -3(-1(1+2) + 2(2-2)) \\ &= -3(-3+0) \\ &= 9. \end{aligned}$$

Since $\det(A) \neq 0$, A is invertible.

When performing the cofactor expansion along the third row of A in Example 6.1.14, we may simply write

$$\det(A) = \begin{vmatrix} 1 & 2 & -1 & 3 \\ 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 3 \\ -1 & 1 & 2 & 1 \end{vmatrix} = -3 \begin{vmatrix} 1 & 2 & -1 \\ 1 & 2 & 0 \\ -1 & 1 & 2 \end{vmatrix}$$

as the other three 3×3 determinants will be multiplied by zero.

Exercise 99

Let $A \in M_{n \times n}(\mathbb{R})$. Show that if A has a row (or column) of zeros, then $\det(A) = 0$.

Section 6.1 Problems

6.1.1. Calculate the determinant of the following matrices.

(a) $\begin{bmatrix} 2 & 1 \\ -3 & 4 \end{bmatrix}$.

(b) $\begin{bmatrix} 4 & 2 \\ 6 & 3 \end{bmatrix}$.

(c) $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 2 & 7 \\ 0 & 0 & 3 \end{bmatrix}$.

(d) $\begin{bmatrix} 1 & 2 & -1 \\ 1 & 2 & 0 \\ 2 & 0 & 0 \end{bmatrix}$.

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

6.1.2. Find all values of $\lambda \in \mathbb{R}$ for which the matrix $A = \begin{bmatrix} 1-\lambda & 2 \\ 2 & 3-\lambda \end{bmatrix}$ is invertible.

6.1.3. (a) Suppose that $A \in M_{n\times n}(\mathbb{R})$ has a column of zeros. Show that $\det(A) = 0$.

(b) Suppose that $A \in M_{n\times n}(\mathbb{R})$ has two identical rows. Show that $\det(A) = 0$.

6.1.4. Find the determinant of the $n \times n$ matrix

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ 3 & 3 & \cdots & 3 \\ \vdots & \vdots & \ddots & \vdots \\ n & n & \cdots & n \end{bmatrix}.$$

6.2 Elementary Row and Column Operations

In Section 6.1, we showed that a matrix $A \in M_{n\times n}(\mathbb{R})$ is invertible if and only if $\det(A) \neq 0$, and we introduced the cofactor expansion as a way of computing $\det(A)$. We noticed in Example 6.1.14 that a matrix with a row (or column) consisting largely of zero entries would lead to a simpler cofactor expansion provided this expansion was performed along that row (or column).

Since Chapter 2, we have been using elementray row operations to carry a matrix to its (reduced) row echelon form. You have likely noticed many zeros are introduced when carrying a matrix to these forms. Hence, it is natural to investigate how elementary row operations affect the determinant of a matrix. In this section, we will see that elementary row operations (and elementary column operations) change the determinant in a predictable way. Thus, with a little "bookkeeping", we will be able to carry a matrix $A \in M_{n\times n}(\mathbb{R})$ to a simpler matrix containing a row or column consisting of mainly zeros. This will allow for easier and faster computation of $\det(A)$.

Example 6.2.1 Consider

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 4 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 4 \\ 1 & 6 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$$

with determinants

$$\det(A) = 2, \quad \det(B) = -2, \quad \det(C) = 2 \quad \text{and} \quad \det(D) = 4.$$

Notice that B, C and D can each be obtained from A by exactly one *elementary column operation*:

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix} \underset{C_1 \leftrightarrow C_2}{\longrightarrow} \begin{bmatrix} 2 & 1 \\ 4 & 1 \end{bmatrix} = B \quad \text{and} \quad \det(B) = -\det(A),$$

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix} \underset{C_2+2C_1 \to C_2}{\longrightarrow} \begin{bmatrix} 1 & 4 \\ 1 & 6 \end{bmatrix} = C \quad \text{and} \quad \det(C) = \det(A),$$

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix} \overset{2C_1 \to C_1}{\longrightarrow} \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix} = D \quad \text{and} \quad \det(D) = 2\det(A).$$

Note that elementary column operations are analogous to elementary row operations. In fact, we may think of performing an elementary column operation to a matrix A as performing the corresponding elementary row operation to A^T.

Recall that for elementary row operations, we write the row operation beside the row that we are modifying (with the exception of row swaps which really modify two rows at once, both of which are clear from our notation). For column operations, we cannot write the column operation "next to" the column we are modifying, so we specify which column we are modifying when writing the operation as done in Example 6.2.1 (as with row swaps, column swaps modify two columns, both of which are clear from our notation).

It's worth pointing out that if we are solving a system of linear equations by carrying the augmented matrix of that system to reduced row echelon form, **then we must never use**

elementary column operations. As an example, the system of linear two equations in two variables with augmented matrix

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right]$$

has the unique solution $x_1 = x_2 = 1$. However, if we apply an elementary column operation, say

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right] \underset{C_2+C_1 \to C_2}{\longrightarrow} \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 1 \end{array}\right],$$

then we arrive at the augmented matrix for a system where $x_1 = x_2 = 1$ is no longer a solution.

Exercise 100 Consider

$$A = \begin{bmatrix} 2 & -1 \\ 6 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 3 \\ 2 & -1 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & -1 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} -6 & 3 \\ 6 & 3 \end{bmatrix}.$$

Show that each of B, C and D can be obtained from A by exactly one elementary row operation, and express each of $\det(B)$, $\det(C)$ and $\det(D)$ in terms of $\det(A)$.

Example 6.2.1 and Exercise 100 suggest that the determinant behaves predictably under elementary row and column operations. The next theorem, stated without proof, shows that this is indeed true.

Theorem 6.2.2 Let $A, B \in M_{n\times n}(\mathbb{R})$.

(a) If B is obtained from A by swapping two distinct rows (or two distinct columns), then $\det(B) = -\det(A)$.

(b) If B is obtained from A by adding a multiple of one row to another row (or a multiple of one column to another column) then $\det(B) = \det(A)$.

(c) If B is obtained from A by multiplying a row (or a column) by $c \in \mathbb{R}$, then $\det(B) = c\det(A)$.

It is important to remember that we never perform elementary row operations and elementary column operations at the same time. In particular, do not add a multiple of a row to a column, or swap a row with a column. If both row and column operations are necessary, then the row operations should be performed in one step and the column operations performed in another.

We now use elementary row and column operations to simplify the computation of determinants.

Example 6.2.3 Find $\det(A)$ if $A = \left[\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{smallmatrix}\right]$.

Solution: Rather than immediately evaluating a cofactor expansion, we will perform elementary row operations to A to introduce two zeros in the first column, and then do a cofactor expansion along that column:

$$\det(A) = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} \underset{\substack{R_2 - 4R_1 \\ R_3 - 7R_1}}{=} \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{vmatrix} = 1 \begin{vmatrix} -3 & -6 \\ -6 & -11 \end{vmatrix}.$$

Of course, we could now evaluate the 2×2 determinant, but for the sake of another example, we will instead multiply the first column by a factor of $-1/3$ and then evaluate the simplified determinant.

$$\det(A) = \begin{vmatrix} -3 & -6 \\ -6 & -11 \end{vmatrix} \overset{-\frac{1}{3}C_1 \to C_1}{=} (-3) \begin{vmatrix} 1 & -6 \\ 2 & -11 \end{vmatrix} = (-3)(-11 + 12) = -3.$$

We make a couple of notes regarding Example 6.2.3. First, we are using "=" rather than "$\longrightarrow$" when we perform our elementary operations on A. This is because we are really working with determinants, and provided we are making the necessary adjustments mentioned in Theorem 6.2.2, we will maintain equality. Secondly, when we performed the operation $-\frac{1}{3}C_1 \to C_1$, a factor of -3 appeared in front of the resulting determinant rather than a factor of $-1/3$. To see why this is, consider

$$C = \begin{bmatrix} -3 & -6 \\ -6 & -11 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & -6 \\ 2 & -11 \end{bmatrix}.$$

Since

$$C = \begin{bmatrix} -3 & -6 \\ -6 & -11 \end{bmatrix} \overset{-\frac{1}{3}C_1 \to C_1}{\longrightarrow} \begin{bmatrix} 1 & -6 \\ 2 & -11 \end{bmatrix} = B,$$

we see that B is obtained from C by multiplying the first column of C by $-1/3$. Thus by Theorem 6.2.2,

$$\det(B) = -\frac{1}{3}\det(C)$$

and so

$$\det(C) = -3\det(B),$$

which is why we have

$$\begin{bmatrix} -3 & -6 \\ -6 & -11 \end{bmatrix} = -3 \begin{bmatrix} 1 & -6 \\ 2 & -11 \end{bmatrix}.$$

We normally view this type of row or column operation as "factoring out" of that row or column, and we omit writing this type of operation as we reduce.

Example 6.2.4 Let $A = \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$. Show that $\det(A) = (b-a)(c-a)(c-b)$.

Solution: We again introduce two zeros into the first column by performing elementary row operations on A, and then do a cofactor expansion along that column. We have

$$\begin{aligned}
\det(A) = \begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} \underset{\substack{R_2-R_1 \\ R_3-R_1}}{=} \begin{vmatrix} 1 & a & a^2 \\ 0 & b-a & b^2-a^2 \\ 0 & c-a & c^2-a^2 \end{vmatrix} &= 1 \begin{vmatrix} (b-a) & (b-a)(b+a) \\ (c-a) & (c-a)(c+a) \end{vmatrix} \\
&= (b-a)(c-a) \begin{vmatrix} 1 & b+a \\ 1 & c+a \end{vmatrix} \\
&= (b-a)(c-a)(c+a-b-a) \\
&= (b-a)(c-a)(c-b).
\end{aligned}$$

In Example 6.2.4, note that the equality

$$\begin{vmatrix} (b-a) & (b-a)(b+a) \\ (c-a) & (c-a)(c+a) \end{vmatrix} = (b-a)(c-a) \begin{vmatrix} 1 & b+a \\ 1 & c+a \end{vmatrix} \tag{6.1}$$

results from factoring out $b-a$ from the first row of the determinant on the left and factoring out $c-a$ from the second row. This corresponds to the row operations $\frac{1}{b-a}R_1 \to R_1$ and $\frac{1}{c-a}R_2 \to R_2$. It is natural to ask what happens if $a = b$ or $a = c$, since it would appear that we are dividing by zero in these cases. However, if $a = b$ or $a = c$, we see that both sides of (6.1) evaluate to zero, so that we still have equality.

Example 6.2.5

Let $B = \begin{bmatrix} 1-\lambda & -2 & 1 \\ 2 & 3-\lambda & 2 \\ -2 & -4 & -3-\lambda \end{bmatrix}$. For what values of $\lambda \in \mathbb{R}$ is $\det(B) = 0$?

Solution: We have

$$\det(B) = \begin{vmatrix} 1-\lambda & -2 & 1 \\ 2 & 3-\lambda & 2 \\ -2 & -4 & -3-\lambda \end{vmatrix} \underset{R_3+R_2}{=} \begin{vmatrix} 1-\lambda & -2 & 1 \\ 2 & 3-\lambda & 2 \\ 0 & -1-\lambda & -1-\lambda \end{vmatrix} \underset{C_2-C_3\to C2}{=} \begin{vmatrix} 1-\lambda & -3 & 1 \\ 2 & 1-\lambda & 2 \\ 0 & 0 & -1-\lambda \end{vmatrix}.$$

Performing a cofactor expansion along the third row gives

$$\det(B) = (-1-\lambda) \begin{vmatrix} 1-\lambda & -3 \\ 2 & 1-\lambda \end{vmatrix} = (-1-\lambda)\big((1-\lambda)^2 + 6\big).$$

Note that $(1-\lambda)^2 + 6 > 0$, so $\det(B) = 0$ implies that $-1-\lambda = 0$, that is, $\lambda = -1$.

Exercise 101

Consider $A = \begin{bmatrix} x & x & 1 \\ x & 1 & x \\ 1 & x & x \end{bmatrix}$. For what values of $x \in \mathbb{R}$ is A not invertible?

Example 6.2.6

Compute $\det(A)$ if $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{bmatrix}$.

Solution:

$$\det(A) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{vmatrix} = 1 \begin{vmatrix} 3 & 0 & 0 \\ 5 & 6 & 0 \\ 8 & 9 & 10 \end{vmatrix} = 1(3) \begin{vmatrix} 6 & 0 \\ 9 & 10 \end{vmatrix} = 1(3)(6)(10) = 180.$$

Note that in the previous example, $\det(A)$ is just the product of the entries on the main diagonal.[1]

Definition 6.2.7

Upper and Lower Triangular Matrices, Diagonal Matrices

Let $A \in M_{n\times n}(\mathbb{R})$. A is called **upper triangular** if every entry below the main diagonal is zero, and A is called **lower triangular** if every entry above the main diagonal is zero. A is called **diagonal** if it is both upper triangular and lower triangular.

Example 6.2.8

The matrices

$$\begin{bmatrix} 1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 6 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 10 \\ 0 & 0 & -2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

are upper triangular, and the matrices

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

are lower triangular. The matrices

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & -5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

are diagonal.

In particular, the $n \times n$ identity matrix I_n and the $n \times n$ zero matrix $0_{n\times n}$ are diagonal matrices.

As evidenced in Example 6.2.6, we have the following result which we state without proof.

[1] Recall that for $A = [a_{ij}] \in M_{n\times n}(\mathbb{R})$, the main diagonal of A consists of the entries $a_{11}, a_{22}, \ldots, a_{nn}$.

Theorem 6.2.9 If $A = [a_{ij}] \in M_{n\times n}(\mathbb{R})$ is a upper or lower triangular, then

$$\det A = a_{11}a_{22}\cdots a_{nn} = \prod_{i=1}^{n} a_{ii}.$$

Note that since a diagonal matrix is upper and lower triangular, Theorem 6.2.9 holds for diagonal matrices as well.

Example 6.2.10 Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & -1 & 2 \\ 3 & -3 & 0 \\ 5 & 0 & 0 \end{bmatrix}.$$

Compute $\det(A)$, $\det(B)$ and $\det(C)$.

Solution: Since A is diagonal, we can apply Theorem 6.2.9 to obtain $\det(A) = 1^3 = 1$.

We present three ways to determine $\det(B)$. We can apply Theorem 6.1.11 to conclude that since B is not invertible, $\det(B) = 0$. We can also use the fact that B has a row (or column) of zeros to conclude that $\det(B) = 0$ by Theorem 6.2.2. Since B is a diagonal matrix, we can also apply Theorem 6.2.9 to arrive at $\det(B) = 0^2 = 0$.

To compute $\det(C)$, we apply elementary row operations to C to obtain a lower triangular matrix. We have

$$\det(C) = \begin{vmatrix} 1 & -1 & 2 \\ 3 & -3 & 0 \\ 5 & 0 & 0 \end{vmatrix} \underset{R_1 \leftrightarrow R_3}{=} (-1)\begin{vmatrix} 5 & 0 & 0 \\ 3 & -3 & 0 \\ 1 & -1 & 2 \end{vmatrix} = (-1)(5)(-3)(2) = 30.$$

Note that in Example 6.2.10, $A = I_3$ and $B = 0_{2\times 2}$. We can similarly show that for every $n \geq 1$,

$$\det(I_n) = 1^n = 1 \quad \text{and} \quad \det(0_{n\times n}) = 0^n = 0.$$

Example 6.2.11 Let $A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 5 & 6 & 7 \end{bmatrix}$. Compute $\det(A)$ by using elementary row operations to carry A to an upper triangular matrix.

Solution: We have

$$\det(A) = \begin{vmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 5 & 6 & 7 \end{vmatrix} \underset{\substack{R_2 - \frac{3}{2}R_1 \\ R_3 - \frac{5}{2}R_1}}{=} \begin{vmatrix} 2 & 3 & 4 \\ 0 & -1/2 & -1 \\ 0 & -3/2 & -3 \end{vmatrix} \underset{R_3 - 3R_2}{=} \begin{vmatrix} 2 & 3 & 4 \\ 0 & -1/2 & -1 \\ 0 & 0 & 0 \end{vmatrix},$$

so

$$\det(A) = 2\left(-\frac{1}{2}\right)(0) = 0.$$

Exercise 102 Let $A = \begin{bmatrix} -1 & 4 & 3 \\ 2 & 0 & -2 \\ 2 & 3 & -2 \end{bmatrix}$. Compute $\det(A)$ by using elementary column operations to carry A to a lower triangular matrix.

Section 6.2 Problems

6.2.1. Find the determinant of the following matrices.

(a) $\begin{bmatrix} 1 & 2 & -1 \\ 1 & 2 & -1 \\ 2 & 4 & 3 \end{bmatrix}$.

(b) $\begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -5 \\ 6 & 3 & 2 \end{bmatrix}$.

(c) $\begin{bmatrix} 2 & -2 & 0 & 1 \\ 1 & 3 & 0 & -1 \\ -2 & 2 & 0 & -3 \\ 6 & 1 & 2 & 0 \end{bmatrix}$.

6.2.2. Let $x \in \mathbb{R}$, and let

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & x & x \\ 1 & x & 0 & x \\ 1 & x & x & 0 \end{bmatrix}.$$

(a) Express $\det(A)$ in terms of x.

(b) Determine all values of $x \in \mathbb{R}$ for which A is invertible.

6.2.3. If $\begin{vmatrix} a & b & c \\ p & q & r \\ x & y & z \end{vmatrix} = -7$, find $\begin{vmatrix} 2b & 2a & 2c \\ q+b & p+a & r+c \\ y & x & z \end{vmatrix}$.

6.3 Properties of Determinants

In this section, we explore the algebraic properties of the determinant. We will see that the determinant behaves well with respect to scalar multiplication and matrix multiplication, but not with matrix addition. We first examine how the determinant behaves with respect to scalar multiplication.

Example 6.3.1 Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. Express $\det(2A)$ in terms of $\det(A)$.

Solution: To obtain A from $2A$, we perform elementary row operations. Specifically, we multiply each row of $2A$ by $\frac{1}{2}$. Performing these row operations one at a time, we have

$$\det(2A) = \begin{vmatrix} 2 & 0 & 2 \\ 0 & 2 & 2 \\ 2 & 2 & 0 \end{vmatrix} = 2\begin{vmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 2 & 2 & 0 \end{vmatrix} = 2^2\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 2 & 2 & 0 \end{vmatrix} = 2^3\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2^3\det(A).$$

Example 6.3.1 appears to indicate that if a matrix A is multiplied by a scalar c, then the resulting determinant is scaled by a factor c^n, where n is the number of rows (or columns) in A. This is verified in the following theorem.

Theorem 6.3.2 If $A \in M_{n\times n}(\mathbb{R})$ and $c \in \mathbb{R}$, then $\det(cA) = c^n \det(A)$.

Proof: If $c = 0$, then $\det(cA) = \det(0_{n\times n}) = 0$ and $c^n \det(A) = 0^n(\det(A)) = 0$ so the result holds. If $c \neq 0$, then we perform $\frac{1}{c}R_i \to R_i$ to each of the n rows of cA, which gives the result by Theorem 6.2.2. □

Exercise 103 Let $A \in M_{n\times n}(\mathbb{R})$. Suppose that $\det(-2A) = 64\det(A)$. Determine n.

Next, we investigate how the determinant behaves with respect to matrix multiplication.

Example 6.3.3 Find $\det(A)\det(B)$ and $\det(AB)$ where $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$.

Solution: We have

$$\det(A)\det(B) = \big(4-6\big)\big(2-(-1)\big) = -2(3) = -6,$$

and

$$\det(AB) = \begin{vmatrix} -1 & 5 \\ -1 & 11 \end{vmatrix} = -11-(-5) = -6.$$

Example 6.3.3 illustrates a general phenomenon, which we state formally in the next theorem.

Theorem 6.3.4 If $A, B \in M_{n\times n}(\mathbb{R})$, then $\det(AB) = \det(A)\det(B)$.

Theorem 6.3.4 says that for $n \times n$ matrices, the determinant distributes over matrix multiplication. Since multiplication of real numbers is commutative, we have

$$\det(AB) = \det(A)\det(B) = \det(B)\det(A) = \det(BA)$$

for any $A, B \in M_{n\times n}(\mathbb{R})$. This means that even though A and B do not commute in general, we are guaranteed that $\det(AB) = \det(BA)$.

Theorem 6.3.4 generalizes to more than two matrices. For $A_1, A_2, \ldots, A_k \in M_{n\times n}(\mathbb{R})$, we have

$$\det(A_1A_2\cdots A_k) = \det(A_1)\det(A_2)\cdots\det(A_k).$$

In particular, if $A_1 = A_2 = \cdots = A_k = A$ for any positive integer k, then we obtain

$$\det(A^k) = (\det(A))^k.$$

Example 6.3.5 Let $A = \begin{bmatrix} 3 & 5 \\ 1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 2 \\ -1 & -3 \end{bmatrix}$. Compute $\det(AB)$ and $\det(A^4)$.

Solution: We have

$$\det(A) = \begin{vmatrix} 3 & 5 \\ 1 & 3 \end{vmatrix} = 9 - 5 = 4,$$

and

$$\det(B) = \begin{vmatrix} 3 & 2 \\ -1 & -3 \end{vmatrix} = -9 - (-2) = -7.$$

It follows from Theorem 6.3.4 that

$$\det(AB) = \det(A)\det(B) = 4(-7) = -28,$$

and

$$\det(A^4) = (\det(A))^4 = 4^4 = 256.$$

Recalling the generalization of Theorem 3.5.6(b), we have that if $A_1, A_2, \ldots, A_k \in M_{n\times n}(\mathbb{R})$ are invertible, then the product $A_1A_2\cdots A_k$ is invertible and

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}\cdots A_2^{-1}A_1^{-1}.$$

The next example shows that if a product of $n \times n$ matrices is invertible, then each matrix in the product is invertible.

Example 6.3.6 Let $A_1, A_2, \ldots, A_k \in M_{n\times n}(\mathbb{R})$ be such that the product $A_1A_2\cdots A_k$ is invertible. Then by Theorem 6.3.4,

$$0 \neq \det(A_1A_2\cdots A_k) = \det(A_1)\det(A_2)\cdots\det(A_k).$$

Thus for $i = 1, 2, \ldots, k$, we have that $\det(A_i) \neq 0$ and thus A_i is invertible for $i = 1, \ldots, k$.

We now use Theorem 6.3.4 to compute the determinant of the inverse of a matrix.

Theorem 6.3.7 Let $A \in M_{n\times n}(\mathbb{R})$ be invertible. Then

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Proof: Let $A \in M_{n\times n}(\mathbb{R})$ be an invertible matrix. By Theorem 6.3.4, we have

$$\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det I = 1.$$

Since A invertible implies $\det(A) \neq 0$, we obtain

$$\det(A^{-1}) = \frac{1}{\det(A)}. \qquad \square$$

For an invertible matrix $A \in M_{n\times n}(\mathbb{R})$, we define $A^{-k} = (A^{-1})^k$ for any positive integer k and we define $A^0 = I$. Thus

$$\det(A^{-k}) = \det\left((A^{-1})^k\right) = \left(\det(A^{-1})\right)^k = \left((\det(A))^{-1}\right)^k = (\det(A))^{-k}$$

and

$$\det(A^0) = \det(I) = 1 = (\det(A))^0.$$

It follows that

$$\det(A^k) = (\det(A))^k$$

for any integer k where $k \leq 0$ requires that A be invertible.

Example 6.3.8 Let $A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \\ -4 & -2 & -5 \end{bmatrix}$. Find $\det(A^{-5})$.

Solution: We have

$$\det(A) = \begin{vmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \\ -4 & -2 & -5 \end{vmatrix} \underset{\substack{R_2 - \frac{1}{2}R_1 \\ R_3 + 2R_1}}{=} \begin{vmatrix} 2 & 1 & 3 \\ 0 & 3/2 & -1/2 \\ 0 & 0 & 1 \end{vmatrix}.$$

Thus

$$\det(A) = 2\left(\frac{3}{2}\right)(1) = 3,$$

and so

$$\det(A^{-5}) = (\det(A))^{-5} = 3^{-5} = \frac{1}{243}.$$

We now look at an example involving the determinant of a square matrix and its transpose.

Example 6.3.9 Let $A = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 3 & 0 \\ 1 & 2 & 1 \end{bmatrix}$. Compute $\det(A)$ and $\det(A^T)$.

Solution: Performing a cofactor expansion along the third column of A gives

$$\det(A) = \begin{vmatrix} 1 & 1 & 2 \\ -1 & 3 & 0 \\ 1 & 2 & 1 \end{vmatrix} = 2 \begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix} + 1 \begin{vmatrix} 1 & 1 \\ -1 & 3 \end{vmatrix} = 2(-2-3) + 1(3+1) = -6.$$

We compute

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

and performing a cofactor expansion along the third row of A^T gives

$$\det(A^T) = \begin{vmatrix} 1 & -1 & 1 \\ 1 & 3 & 2 \\ 2 & 0 & 1 \end{vmatrix} = 2 \begin{vmatrix} -1 & 1 \\ 3 & 2 \end{vmatrix} + 1 \begin{vmatrix} 1 & -1 \\ 1 & 3 \end{vmatrix} = 2(-2-3) + 1(3+1) = -6.$$

Example 6.3.9 supports the idea that $\det(A^T) = \det(A)$ for $A \in M_{n\times n}(\mathbb{R})$. This is indeed true, as stated in the next theorem.

Theorem 6.3.10 Let $A \in M_{n\times n}(\mathbb{R})$. Then $\det(A^T) = \det(A)$.

We do not prove Theorem 6.3.10, but the next exercise hints at one way this could be proven (although there is a better way to prove this that is beyond the scope of this course).

Exercise 104 Consider A from Example 6.3.9.

(a) Compute $\det(A)$ by using elementary row operations to carry A to an upper triangular matrix.

(b) Compute $\det(A^T)$ by using elementary column operations to carry A^T to a lower triangular matrix.

(c) How are the column operations used in part (b) related to the row operations used in part (a)?

The next example combines many of the results discussed in this section.

Example 6.3.11 If $\det(A) = 3$, $\det(B) = -2$ and $\det(C) = 4$ for $A, B, C \in M_{n\times n}(\mathbb{R})$, find

$$\det(A^2 B^T C^{-1} B^2 (A^{-1})^2).$$

Solution: We have

$$\begin{aligned}\det(A^2B^TC^{-1}B^2(A^{-1})^2) &= \det(A^2)\det(B^T)\det(C^{-1})\det(B^2)\det((A^{-1})^2)\\ &= (\det A)^2(\det B)\frac{1}{\det C}(\det B)^2\frac{1}{(\det A)^2}\\ &= \frac{(\det B)^3}{\det C}\\ &= \frac{(-2)^3}{4} = -\frac{8}{4} = -2.\end{aligned}$$

Finally, we turn to matrix addition. As the next example shows, the determinant does not behave well with matrix addition.

Example 6.3.12 Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $B \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Then

$$\det(A) + \det(B) = 0 + 0 = 0$$

but

$$\det(A + B) = \det(I) = 1,$$

showing that $\det(A + B) \neq \det(A) + \det(B)$.

Exercise 105 Find two nonzero matrices, $A, B \in M_{2\times 2}(\mathbb{R})$, such that $\det(A + B) = \det(A) + \det(B)$.

Section 6.3 Problems

6.3.1. Let A, B and C be $n \times n$ matrices with $\det(A) = 1$, $\det(B) = -3$ and $\det(C) = 4$. Compute $\det\left(A^2 B^T C^{-1} A^{-1} B^2\right)$.

6.3.2. (a) Let $A \in M_{n\times n}(\mathbb{R})$ be such that $A = -A^T$. Prove that if n is odd, then $\det(A) = 0$.

(b) Let $P \in M_{n\times n}(\mathbb{R})$ be such that $P^2 = uP$ for some real number $u \neq 0$. Find all possible values of $\det(P)$.

6.3.3. Let $A, B \in M_{n\times n}(\mathbb{R})$. Prove that if AB^T is invertible, then A and B are invertible.

6.4 Optional Section: Area, Volume and Determinants

The determinant of a matrix $A \in M_{n\times n}(\mathbb{R})$ was introduced in Section 6.1 as a number that indicates if A is invertible or not. Thus, our focus has been on whether or not the determinant of A is zero or nonzero. In this section, we will see that the determinant of A has a very nice geometric meaning as well: it can be interpreted as the area of a parallelogram or the volume of a parallelepiped (the 3-dimensional version of a parallelogram). We will extend this idea to see how linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$ or $T : \mathbb{R}^3 \to \mathbb{R}^3$ change the volume of common shapes like circles and spheres in a predictable way.

Theorem 1.8.3 allows us to use the cross product to find the area of a parallelogram determined by two vectors in $\mathbb{R}^3$. We now consider the problem of finding the area of a parallelogram determined by two vectors in $\mathbb{R}^2$. Although Theorem 1.8.3 is only valid for vectors in $\mathbb{R}^3$, we will see that it can be used to prove the following result.

Theorem 6.4.1 **(Area of a Parallelogram in $\mathbb{R}^2$)**

The area of the parallelogram, P, determined by $\vec{x}, \vec{y} \in \mathbb{R}^2$ is given by

$$\text{area}(P) = \left|\det\left(\begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}\right)\right|.$$

Proof: Let $\vec{x}, \vec{y} \in \mathbb{R}^2$ and $\vec{x}_0, \vec{y}_0 \in \mathbb{R}^3$ with

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \vec{x}_0 = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{y}_0 = \begin{bmatrix} y_1 \\ y_2 \\ 0 \end{bmatrix}.$$

Let $P \subseteq \mathbb{R}^2$ be the parallelogram determined by $\vec{x}$ and $\vec{y}$, and let $P_0 \subseteq \mathbb{R}^3$ be the parallelogram determined by $\vec{x}_0$ and $\vec{y}_0$ as shown in Figure 6.4.1.

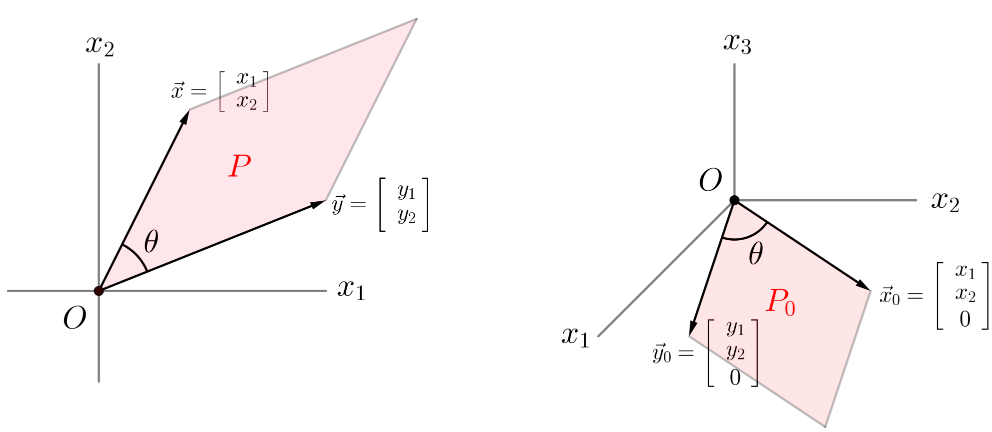


Figure 6.4.1: A parallelogram P determined by $\vec{x}, \vec{y} \in \mathbb{R}^2$ on the left, and its "realization" P_0 lying in the x_1x_2-plane of $\mathbb{R}^3$ on the right.

We see that $\|\vec{x}\| = \|\vec{x}_0\|$, $\|\vec{y}\| = \|\vec{y}_0\|$, and that $\vec{x} \cdot \vec{y} = \vec{x}_0 \cdot \vec{y}_0$. From this, it follows that

$$\text{area}(P) = \text{area}(P_0).$$

Thus by Theorem 1.8.3,

$$\begin{aligned}\text{area}(P) &= \left\| \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0 \\ 0 \\ x_1y_2 - y_1x_2 \end{bmatrix} \right\| = \sqrt{(x_1y_2 - y_1x_2)^2} \\ &= |x_1y_2 - y_1x_2| = \left| \det\left(\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \right) \right| = \left| \det\left(\begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix} \right) \right|. \qquad \square\end{aligned}$$

Example 6.4.2 Let P be the parallelogram determined by $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Find the area of P.

Solution: By Theorem 6.4.1, the area of P is

$$\text{area}(P) = \left| \det \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \right| = |4 - 6| = |-2| = 2.$$

We make a note here about notation. For

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \in M_{n\times n}(\mathbb{R}),$$

we have previously introduced the notation

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

to denote $\det(A)$. However, when talking about the *absolute value* of $\det(A)$, we **must not write**

$$\left\| \begin{matrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{matrix} \right\|$$

to denote $\big|\det(A)\big|$ since this has a different meaning in linear algebra.[2] Instead, we must write

$$\left| \det \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \right|$$

to when denoting $\big|\det(A)\big|$.

Let $\vec{x}, \vec{y} \in \mathbb{R}^2$ and let P denote the parallelogram they determine. For a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$, we denote the parallelogram determined by $T(\vec{x})$ and $T(\vec{y})$ by $T(P)$. See Figure 6.4.2.

[2]For a matrix $A \in M_{m\times n}(\mathbb{R})$, the symbol $\|A\|$ denotes a *matrix norm* of A, which is studied in later linear algebra courses.

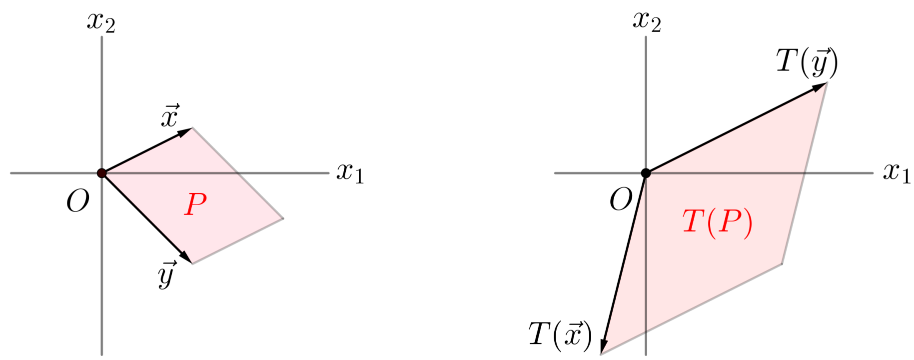


Figure 6.4.2: The parallelogram $P \subseteq \mathbb{R}^2$ determined by $\vec{x}$ and $\vec{y}$, and the parallelogram $T(P) \subseteq \mathbb{R}^2$ determined by $T(\vec{x})$ and $T(\vec{y})$.

From Theorem 6.4.1 the area of $T(P)$ is given by

$$\operatorname{area}(T(P)) = \left|\det\left(\begin{bmatrix} T(\vec{x}) & T(\vec{y}) \end{bmatrix}\right)\right|.$$

However, the next theorem shows that we can obtain a more meaningful formula for $\operatorname{area}(T(P))$ using Theorem 6.3.4.

Theorem 6.4.3 Let $\vec{x}, \vec{y} \in \mathbb{R}^2$ and let P be the parallelogram they determine. For a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$, the area of the parallelogram, $T(P)$, determined by $T(\vec{x})$ and $T(\vec{y})$ is given by

$$\operatorname{area}(T(P)) = \left|\det\left(\begin{bmatrix} T \end{bmatrix}\right)\right| \operatorname{area}(P).$$

Proof: We have

$$\begin{aligned}
\operatorname{area}(T(P)) &= \left|\det\left(\begin{bmatrix} T(\vec{x}) & T(\vec{y}) \end{bmatrix}\right)\right| && \text{by Theorem 6.4.1} \\
&= \left|\det\left(\begin{bmatrix} [T]\,\vec{x} & [T]\,\vec{y} \end{bmatrix}\right)\right| \\
&= \left|\det\left(\begin{bmatrix} T \end{bmatrix} \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}\right)\right| && \text{by Definition 3.4.1} \\
&= \left|\det\left(\begin{bmatrix} T \end{bmatrix}\right) \det\left(\begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}\right)\right| && \text{by Theorem 6.3.4} \\
&= \left|\det\left(\begin{bmatrix} T \end{bmatrix}\right)\right| \left|\det\left(\begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}\right)\right| \\
&= \left|\det\left(\begin{bmatrix} T \end{bmatrix}\right)\right| \operatorname{area}(P) && \text{by Theorem 6.4.1.}
\end{aligned}$$ □

What is interesting about the result of Theorem 6.4.3 is that it does not depend explicitly on the vectors $\vec{x}$ and $\vec{y}$ that determine P. Theorem 6.4.3 shows that if we have a parallelogram $P \subseteq \mathbb{R}^2$ and we apply a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$, then the area of P will be scaled by a factor of $\left|\det\left(\begin{bmatrix} T \end{bmatrix}\right)\right|$ under T.

Example 6.4.4 Let P be a parallelogram with $\operatorname{area}(P) = 4$. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with standard matrix

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 1 & 1 \end{bmatrix}.$$

Determine $\operatorname{area}(T(P))$.

Solution: Using Theorem 6.4.3, the area of $T(P)$ is given by

$$\text{area}(T(P)) = \left|\det\left(\left[\,T\,\right]\right)\right| \text{area}(P) = |-4|(4) = 4(4) = 16.$$

Exercise 106 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a horizontal shear by $s > 0$ and let P be a parallelogram with area$(P) = a$ where $a \geq 0$. Determine area$(T(P))$.

Although stated for parallelograms, Theorem 6.4.3 generalizes to many shapes in $\mathbb{R}^2$, such as circles, ellipses and polygons.

Example 6.4.5 Consider a circle C of radius $r = 1$ centred at the origin in $\mathbb{R}^2$. The area of this circle is

$$\text{area}(C) = \pi r^2 = \pi(1)^2 = \pi.$$

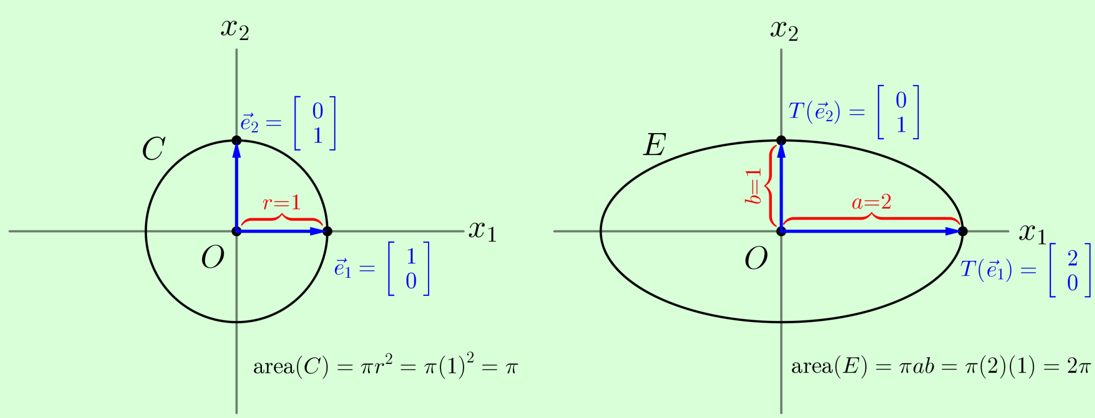
If we consider a stretch in the x_1-direction by a factor of 2, then we are considering the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with standard matrix

$$\left[\,T\,\right] = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

We denote image of our circle under T by $E = T(C)$, which is an ellipse. By the generalized version of Theorem 6.4.3 mentioned above, this ellipse has area

$$\text{area}(E) = \left|\det\left(\left[\,T\,\right]\right)\right| \text{area}(C) = |2|\pi = 2\pi.$$

The following figure depicts our circle along with the resulting ellipse, and shows that our result for the area of the ellipse is consistent with the actual formula for the area of an ellipse.



Note that our choice of C being centred at the origin was arbitrary - we would obtain the same result for any circle of radius 1 (but the above figure is easier to digest if C is centered at the origin!).

Exercise 107 A polygon Q has area$(Q) = 2$. Find the area of $T(Q)$ if $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a vertical shear by a factor of 3, followed by a contraction by a factor of $\frac{1}{2}$.

We now turn our attention to considering volumes. Recall that three vectors $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ determine a parallelpiped Q (see Figure 1.8.2) and that the volume of Q is given by

$$\text{vol}(Q) = |\vec{z} \cdot (\vec{x} \times \vec{y})|,$$

as was verified in Theorem 1.8.5. Analogous to Theorem 6.4.1, we can use determinants to compute the volume of parallelepipeds in $\mathbb{R}^3$.

Theorem 6.4.6 **(Volume of a Parallelepiped in $\mathbb{R}^3$)**

The volume of the parallelepiped, Q, determined by $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ is given by

$$\text{vol}(Q) = \left|\det\left(\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right)\right|.$$

Proof: Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

Then

$$\begin{aligned}
\text{vol}(Q) = |\vec{z} \cdot (\vec{x} \times \vec{y})| &= \left| \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \cdot \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right) \right| = \left| \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \cdot \begin{bmatrix} x_2y_3 - y_2x_3 \\ -(x_1y_3 - y_1x_3) \\ x_1y_2 - y_1x_2 \end{bmatrix} \right| \\
&= |z_1(x_2y_3 - y_2x_3) - z_2(x_1y_3 - y_1x_3) + z_3(x_1y_2 - y_1x_2)| \\
&= \left| z_1 \det\left(\begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix}\right) - z_2 \det\left(\begin{bmatrix} x_1 & y_1 \\ x_3 & y_3 \end{bmatrix}\right) + z_3 \det\left(\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}\right) \right| \\
&= \left| \det\left(\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{bmatrix}\right) \right| = \left|\det\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right|.
\end{aligned}$$

□

Example 6.4.7 Let Q be the parallelepiped determined by $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, $\vec{z} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$. Compute vol$(Q)$.

Solution: By Theorem 6.4.6, we have

$$\begin{aligned}
\text{vol}(Q) = \left|\det\left(\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right)\right| &= \left| \det\left(\begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix}\right) \right| \\
&= \left| 1\det\left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\right) - 2\det\left(\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}\right) - 1\det\left(\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}\right) \right| \\
&= |1(1) - 2(5) - 1(3)| \\
&= |-12| \\
&= 12.
\end{aligned}$$

Exercise 108 Let Q be the parallelpiped determined by $\vec{x} = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\vec{z} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}$. Determine $\mathrm{vol}(Q)$.

We also have the following:

Theorem 6.4.8 Let Q be the parallelepiped determined by $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$. For a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$, the volume of the parallepiped, $T(Q)$, determined by $T(\vec{x})$, $T(\vec{y})$ and $T(\vec{z})$ is given by

$$\mathrm{vol}(T(Q)) = \left|\det\left(\left[\,T\,\right]\right)\right| \mathrm{vol}(Q).$$

Exercise 109 Prove Theorem 6.4.8. Hint: Mimic the proof of Theorem 6.4.3.

Example 6.4.9 Let Q be a parallelepiped with $\mathrm{vol}(Q) = 7$. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation with standard matrix

$$\left[\,T\,\right] = \begin{bmatrix} 5 & 2 & 7 \\ 0 & 0 & -4 \\ 0 & 5 & 9 \end{bmatrix}.$$

Determine $\mathrm{vol}(T(Q))$.

Solution: We compute

$$\det\left(\left[\,T\,\right]\right) = \begin{vmatrix} 5 & 2 & 7 \\ 0 & 0 & -4 \\ 0 & 5 & 9 \end{vmatrix} \underset{R_2 \leftrightarrow R_3}{=} (-1) \begin{vmatrix} 5 & 2 & 7 \\ 0 & 5 & 9 \\ 0 & 0 & -4 \end{vmatrix} = (-1)(5)(5)(-4) = 100$$

Using Theorem 6.4.8, the volume of $T(Q)$ is given by

$$\mathrm{vol}(T(Q)) = \left|\det\left(\left[\,T\,\right]\right)\right| \mathrm{vol}(Q) = |100|(7) = 700.$$

As with Theorem 6.4.3, Theorem 6.4.8 generalizes to many shapes in $\mathbb{R}^3$ other than parallelpipeds.

Example 6.4.10 Consider a sphere S of radius $r = 1$ centred at the origin in $\mathbb{R}^3$. The volume of S is

$$\mathrm{vol}(S) = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi(1)^3 = \frac{4}{3}\pi.$$

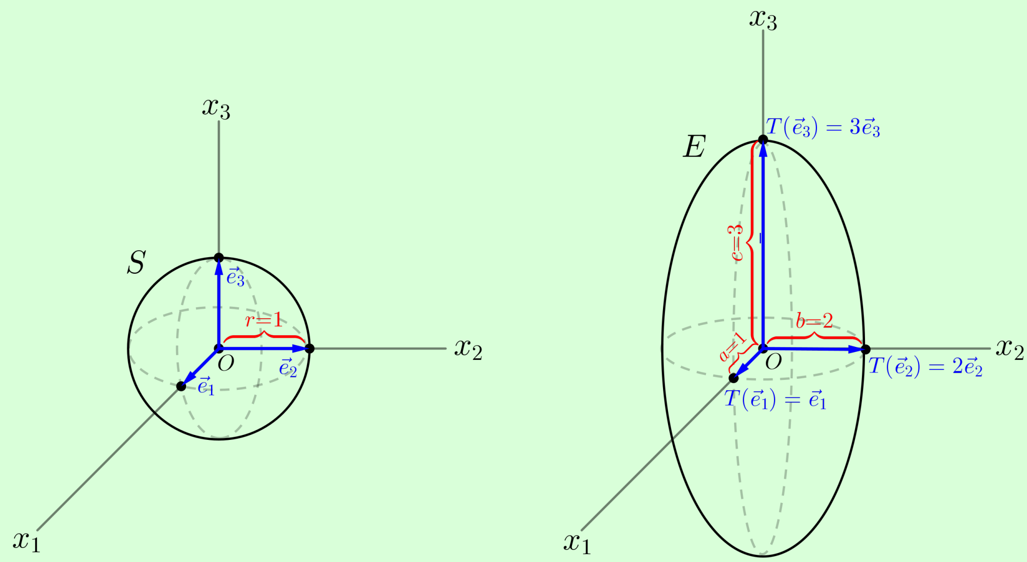
If we consider a stretch in the x_2-direction by a factor of 2 *and* a stretch in the x_3-direction by a factor of 3, then we have the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with standard matrix

$$\left[\,T\,\right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

The image of the sphere S under T is an ellipsoid, E, which we denote by $T(S)$. By the generalized version of Theorem 6.4.8 mentioned above, E has volume

$$\mathrm{vol}(E) = \left|\det\left(\left[T\right]\right)\right| \mathrm{vol}(S) = |6|\frac{4}{3}\pi = 8\pi.$$

The image below illustrates this, and shows that our result for the volume of the ellipsoid is consistent with the actual formula for the volume of an ellipsoid.



$\mathrm{vol}(S) = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi(1)^3 = \frac{4}{3}\pi$

$\mathrm{vol}(E) = \frac{4}{3}\pi abc = \frac{4}{3}\pi(1)(2)(3) = 8\pi$

Section 6.4 Problems

6.4.1. Let Q be the parallelepiped determined by

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Find the volume of Q using

(a) Theorem 6.4.6.

(b) Theorem 1.8.5.

6.4.2. Prove Theorem 6.4.6. [**Hint**: Express $\vec{x}$, $\vec{y}$ and $\vec{z}$ in terms of their components and apply Theorem 1.8.5.]

6.4.3. Let Q be the parallelepiped determined by

$$\vec{v}_1 = \begin{bmatrix} x \\ 0 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ x \end{bmatrix} \quad \text{and} \quad \vec{v}_3 = \begin{bmatrix} 1 \\ x \\ 0 \end{bmatrix}.$$

Determine all values of $x \in \mathbb{R}$ such that $\operatorname{vol}(Q) = 9$.

6.4.4. (a) Let P be the parallelogram determined by

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 4 \\ -3 \end{bmatrix}.$$

(i) Compute $\operatorname{area}(P)$.

(ii) Compute $\sqrt{\det(A^T A)}$ where $A = \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}$. What do you notice?

(b) Let Q be the parallelepiped determined by

$$\vec{x} = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}.$$

(i) Compute $\operatorname{vol}(Q)$.

(ii) Compute $\sqrt{\det(A^T A)}$ where $A = \begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}$. What do you notice?

(c) Let $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^n$ and let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$. Prove that

$$\sqrt{\det(A^T A)} = |\det(A)|.$$

Note: Since

$$|\det(A)| = \left|\det\left(\begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}\right)\right|,$$

the results in this problem gives us the following:

- The area of the parallelogram P determined by $\vec{x}, \vec{y} \in \mathbb{R}^2$ is given by

$$\operatorname{area}(P) = \sqrt{\det(A^T A)}$$

where $A = \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}$.

- The volume of the parallelepiped Q determined by $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ is
$$\operatorname{vol}(Q) = \sqrt{\det(A^T A)}$$
where $A = \begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}$.

6.4.5. (a) Let P be the parallelogram determined by
$$\vec{x} = \begin{bmatrix} 1 \\ 4 \\ -1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}.$$
(i) Compute area(P) using Theorem 1.8.3.
(ii) Compute $\sqrt{\det(A^T A)}$ where $A = \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}$. What do you notice?

(b) Let P be the parallelogram determined by
$$\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 0 \\ 3 \\ 1 \\ 2 \end{bmatrix}.$$
(i) Compute area(P). Do this by defining one of $\vec{x}$ and $\vec{y}$ to represent the base of P and then determine the height of P.
(ii) Compute $\sqrt{\det(A^T A)}$ where $A = \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}$. What do you notice?

(c) Let P be the parallelogram determined by $\vec{x}, \vec{y} \in \mathbb{R}^n$ with $n \geq 2$. Let $A = \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix}$. Prove that
$$\operatorname{area}(P) = \sqrt{\det(A^T A)}.$$
[**Hint**: Show that $\det(A^T A) = \|\vec{x}\|^2 \|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2$. Then show that $(\operatorname{area}(P))^2 = \|\vec{x}\|^2 \|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2$ by following the method used in part (b).]

(d) In part (c), why is it incorrect to say that $\det(A^T A) = \det(A)\det(A^T)$?

(e) In part (c), why is it incorrect use the Lagrange Identity to conclude that $\|\vec{x}\|^2 \|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2 = \|\vec{x} \times \vec{y}\|$?

6.4.6. Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with standard matrix
$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 2 & 3 \\ 1 & 3 & -2 \end{bmatrix}.$$
(a) A region $R \subseteq \mathbb{R}^3$ satisfies $\operatorname{vol}(R) = 6$. Determine $\operatorname{vol}(T(R))$.
(b) A region $S \subseteq \mathbb{R}^3$ satisfies $\operatorname{vol}(T(S)) = 3$. Determine $\operatorname{vol}(S)$.

6.5 Optional Section: Adjugates and Matrix Inverses

In this section we will learn a method that allows us to use determinants to compute the inverse of a matrix. Although this section is optional and will not be covered in class or tested, it does give a very simple way to compute the inverse of a 2×2 matrix that is worth looking at. You are free to use any of the methods developed in this section if you wish.

Recall that for $A = \begin{bmatrix} a \end{bmatrix} \in M_{1\times 1}(\mathbb{R})$, $\det(A) = a$, and A that is invertible if and only if $a \neq 0$. In this case,

$$\begin{bmatrix} a \end{bmatrix} \begin{bmatrix} \frac{1}{a} \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} = I_1$$

so $A^{-1} = \begin{bmatrix} \frac{1}{a} \end{bmatrix}$. Not surprisingly, we can compute the inverse of $A \in M_{1\times 1}(\mathbb{R})$ by inspection.

We now focus our attention on $A \in M_{2\times 2}(\mathbb{R})$.

Example 6.5.1 Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

so that $\det(A) = a_{11}a_{22} - a_{12}a_{21}$. Using cofactor expansions along the rows of A, we know that

$$\begin{aligned} \det(A) &= a_{11}C_{11}(A) + a_{12}C_{12}(A) && \text{(cofactor expansion along the first row of } A\text{)}, \\ &= a_{21}C_{21}(A) + a_{22}C_{22}(A) && \text{(cofactor expansion along the second row of } A\text{)}. \end{aligned}$$

Consider the matrix

$$B = \begin{bmatrix} C_{11}(A) & C_{12}(A) \\ C_{21}(A) & C_{22}(A) \end{bmatrix}^T = \begin{bmatrix} C_{11}(A) & C_{21}(A) \\ C_{12}(A) & C_{22}(A) \end{bmatrix} = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Now

$$\begin{aligned} AB &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{12} & a_{11} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & -a_{11}a_{12} + a_{12}a_{11} \\ a_{21}a_{22} - a_{22}a_{21} & -a_{21}a_{12} + a_{22}a_{11} \end{bmatrix} \\ &= \begin{bmatrix} \det(A) & 0 \\ 0 & \det(A) \end{bmatrix} \\ &= \det(A) I_2. \end{aligned}$$

If A is invertible, then $\det(A) \neq 0$ and it follows from $AB = \det(A)I_2$ that

$$A \left(\frac{1}{\det(A)} B \right) = I_2$$

which shows that

$$A^{-1} = \frac{1}{\det(A)} B.$$

The next exercise asks you to verify a similar property for $A \in M_{3\times 3}(\mathbb{R})$.

Exercise 110 Let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \in M_{3\times 3}(\mathbb{R})$$

and define

$$B = \begin{bmatrix} C_{11}(A) & C_{12}(A) & C_{13}(A) \\ C_{21}(A) & C_{22}(A) & C_{23}(A) \\ C_{31}(A) & C_{32}(A) & C_{33}(A) \end{bmatrix}^T = \begin{bmatrix} C_{11}(A) & C_{21}(A) & C_{31}(A) \\ C_{12}(A) & C_{22}(A) & C_{32}(A) \\ C_{13}(A) & C_{23}(A) & C_{33}(A) \end{bmatrix}.$$

(a) Show that $AB = \det(A)I_3$.

(b) If $\det(A) \neq 0$, give a formula for A^{-1}.

Hint: Part (a) can be quite tedious. You should be able to show that the $(1,1)$-, $(2,2)$- and $(3,3)$-entries of AB are each $\det(A)$. You should also be able to show that a couple of the remaining entries of AB are zero, but don't compute them all as it's quite time consuming.

The matrices B from Example 6.5.1 and Exercise 110 appear to be important. We make the following definition.

Definition 6.5.2
Cofactor Matrix, Adjugate

Let $A \in M_{n\times n}(\mathbb{R})$ with $n \geq 2$.

(a) The **cofactor matrix** of A is

$$\operatorname{cof}(A) = [C_{ij}(A)] \in M_{n\times n}(\mathbb{R}),$$

(b) The **adjugate** of A is

$$\operatorname{adj}(A) = [C_{ij}(A)]^T = \big[C_{ji}(A)\big] \in M_{n\times n}(\mathbb{R}).$$

Recalling Example 6.5.1, we see that we have already computed the adjugate of a 2×2 matrix. For

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

we have

$$\operatorname{cof}(A) = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} \quad \text{and} \quad \operatorname{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Thus we can compute the adjugate of $A \in M_{2\times 2}(\mathbb{R})$ by inspection! We simply swap the main diagonal entries (a and d) and multiply the off-diagonal entries (b and c) by -1.

Example 6.5.3 Compute $\operatorname{adj}(A)$ for $A = \begin{bmatrix} 1 & 3 \\ -2 & 4 \end{bmatrix}$.

Solution: We have
$$\operatorname{adj}(A) = \begin{bmatrix} 4 & -3 \\ 2 & 1 \end{bmatrix}.$$

For $A \in M_{3\times 3}(\mathbb{R})$, we also have a formula for $\operatorname{adj}(A)$. Letting
$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \in M_{3\times 3}(\mathbb{R}),$$
we can use Definition 6.5.2 to compute
$$\operatorname{adj}(A) = \begin{bmatrix} a_{22}a_{33} - a_{23}a_{32} & a_{13}a_{32} - a_{12}a_{33} & a_{12}a_{23} - a_{13}a_{22} \\ a_{23}a_{31} - a_{21}a_{33} & a_{11}a_{33} - a_{13}a_{31} & a_{13}a_{21} - a_{11}a_{23} \\ a_{21}a_{32} - a_{22}a_{31} & a_{12}a_{31} - a_{11}a_{32} & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix}.$$

In fact, given $A = \begin{bmatrix} a_{ij} \end{bmatrix} \in M_{n\times n}(\mathbb{R})$ we can derive similar formulas for $\operatorname{adj}(A)$ for any positive integer n. As one might imagine, such formulas become more complicated for any $n \geq 3$ and are not worth memorizing. It is better in this case to use Definition 6.5.2 and compute the cofactors one-by-one. This is illustrated in the next example.

Example 6.5.4 Compute $\operatorname{adj}(A)$ if $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 3 & 4 & 5 \end{bmatrix}$.

Solution:
$$\operatorname{adj}(A) = \begin{bmatrix} C_{11}(A) & C_{12}(A) & C_{13}(A) \\ C_{21}(A) & C_{22}(A) & C_{23}(A) \\ C_{31}(A) & C_{32}(A) & C_{33}(A) \end{bmatrix}^T = \begin{bmatrix} \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 3 & 5 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ 3 & 4 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 3 & 5 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} \\ \begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} \end{bmatrix}^T$$
$$= \begin{bmatrix} -3 & 1 & 1 \\ 2 & -4 & 2 \\ 1 & 1 & -1 \end{bmatrix}^T = \begin{bmatrix} -3 & 2 & 1 \\ 1 & -4 & 1 \\ 1 & 2 & -1 \end{bmatrix}.$$

Exercise 110 and Example 6.5.4 show that even for $A \in M_{3\times 3}(\mathbb{R})$, computing the adjugate is already an onerous task that is highly error prone. Now consider computing the adjugate of a 4×4 matrix - this would involve computing 16 determinants of 3×3 matrices! When working by hand, one should avoid computing adjugates for anything other than 2×2 matrices.

What we have observed in Example 6.5.1 and Exercise 110 also holds for $A \in M_{n\times n}(\mathbb{R})$ with $n \geq 2$ as is stated in the next theorem. We omit the proof.

Theorem 6.5.5 Let $A \in M_{n\times n}(\mathbb{R})$ with $n \geq 2$. Then

$$A(\mathrm{adj}(A)) = \det(A)I = (\mathrm{adj}(A))A.$$

Moreover, if A is invertible, that is, if $\det(A) \neq 0$, then

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A).$$

The following examples will illustrate that Theorem 6.5.5 is useful for $A \in M_{2\times 2}(\mathbb{R})$, but that it quickly becomes impractical for $A \in M_{n\times n}(\mathbb{R})$ when $n \geq 3$.

Example 6.5.6 Consider $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Compute $\det(A)$, $\mathrm{adj}(A)$ and A^{-1}.

Solution: We compute

$$\det(A) = 1(4) - 2(3) = 4 - 6 = -2$$

and

$$\mathrm{adj}(A) = \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}.$$

Thus by Theorem 6.5.5, we obtain

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A) = \frac{1}{-2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}.$$

Exercise 111 Let $A = \begin{bmatrix} 9 & -7 \\ 7 & -5 \end{bmatrix}$. Compute A^{-1} by

(a) using Theorem 6.5.5,

(b) using the Matrix Inversion Algorithm.

In Example 5.4.6, we used a geometric argument to compute the inverse of the rotation matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

The next example shows this is a straightforward computation using Theorem 6.5.5.

Example 6.5.7 Find the inverse of $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Solution: Since

$$\det(R_\theta) = \cos^2\theta + \sin^2\theta = 1$$

and

$$\operatorname{adj}(R_\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

we see that

$$(R_\theta)^{-1} = \frac{1}{\det(R_\theta)} \operatorname{adj}(R_\theta) = \operatorname{adj}(R_\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Example 6.5.8 Find $\det(A)$, $\operatorname{adj}(A)$ and A^{-1} if $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 4 \\ 1 & 2 & 4 \end{bmatrix}$.

Solution: Using a cofactor expansion along the first row, we obtain

$$\begin{aligned} \det(A) &= 1\begin{vmatrix} 1 & 4 \\ 2 & 4 \end{vmatrix} - 1\begin{vmatrix} 1 & 4 \\ 1 & 4 \end{vmatrix} + 2\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \\ &= 1(4-8) - 1(4-4) + 2(2-1) \\ &= -4 + 2 \\ &= -2 \end{aligned}$$

Then

$$\operatorname{adj}(A) = \begin{bmatrix} \begin{vmatrix} 1 & 4 \\ 2 & 4 \end{vmatrix} & -\begin{vmatrix} 1 & 4 \\ 1 & 4 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \\ -\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 1 & 4 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \\ \begin{vmatrix} 1 & 2 \\ 1 & 4 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 1 & 4 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} \end{bmatrix}^T = \begin{bmatrix} -4 & 0 & 1 \\ 0 & 2 & -1 \\ 2 & -2 & 0 \end{bmatrix}^T = \begin{bmatrix} -4 & 0 & 2 \\ 0 & 2 & -2 \\ 1 & -1 & 0 \end{bmatrix}$$

so

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A) = -\frac{1}{2}\begin{bmatrix} -4 & 0 & 2 \\ 0 & 2 & -2 \\ 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & -1 & 1 \\ -1/2 & 1/2 & 0 \end{bmatrix}.$$

Exercise 112 Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 3 & 4 & 5 \end{bmatrix}$. Compute A^{-1} by

(a) using Theorem 6.5.5,

(b) using the Matrix Inversion Algorithm.

These examples have hopefully convinced you that using Theorem 6.5.5 to compute the inverse of $A \in M_{2\times 2}(\mathbb{R})$ is quite quick and easy, but for $A \in M_{n\times n}(\mathbb{R})$ with $n \geq 3$, the Matrix Inversion Algorithm is the far superior method to compute A^{-1}.

Section 6.5 Problems

6.5.1. For each of

(a) $A = \begin{bmatrix} 3 & -2 \\ 1 & 3 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 4 & -3 & 12 \end{bmatrix}$ (c) $A = \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 1 & 0 & 0 & -1 \end{bmatrix}$

compute A^{-1} by

(i) using Theorem 6.5.5,

(ii) using the Matrix Inversion Algorithm.

6.5.2. Let $A = \begin{bmatrix} 1 & 2 & 14 \\ 6 & -12 & 16 \\ 4 & 2 & 8 \end{bmatrix}$. Given that $\det(A) = 744$, compute the $(3,2)-$entry of A^{-1}.

6.5.3. Let $A \in M_{n\times n}(\mathbb{R})$ be invertible with $n \geq 2$. Prove that $\det\big(\text{adj}(A)\big) = \big(\det(A)\big)^{n-1}$

6.5.4. Let $A \in M_{3\times 3}(\mathbb{R})$ be such that $\det(A) = 2$. Compute $\det\big(9A^{-1} - 3\,\text{adj}(A)\big)$.

Chapter 7

Complex Numbers

7.1 Basic Operations

Recall the number systems you know:

Natural Numbers:	$\mathbb{N} = \{1, 2, 3, \ldots\}$
Integers:	$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
Rational Numbers:	$\mathbb{Q} = \left\{ \frac{a}{b} \;\middle|\; a, b \in \mathbb{Z},\, b \neq 0 \right\}$
Real Numbers:	$\mathbb{R}$, the set or collection of all rational and irrational numbers

Note that every natural number is an integer, every integer is a rational number (with denominator equal to 1) and that every rational number is a real number. That is, we have the containments

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}.$$

Consider the following five equations:

$$x + 3 = 5 \tag{7.1}$$
$$x + 4 = 3 \tag{7.2}$$
$$2x = 1 \tag{7.3}$$
$$x^2 = 2 \tag{7.4}$$
$$x^2 = -2 \tag{7.5}$$

Equation (7.1) has solution $x = 2$, and thus can be solved using natural numbers. Equation (7.2) does not have a solution in the natural numbers, but it does have a solution in the integers, namely $x = -1$. Equation (7.3) does not have a solution in the integers, but it does have a rational solution of $x = \frac{1}{2}$. Equation (7.4) does not have a rational solution, but it does a have real solution: $x = \sqrt{2}$. Finally, since the square of any real number is greater than or equal to zero, Equation (7.5) does not have a real solution. In order to solve this last equation, we will need a "larger" set of numbers.

We do this by introducing an "imaginary" object i that satisfies the equation $i^2 = -1$. We will have to explain the rules of working with such an object. Once we do this, we'll find that we have created a very powerful and useful mathematical structure. Although this might seem strange at first sight, it really is not that much different from introducing a number such as $x = \sqrt{2}$, which – if you really think about it – is nothing other than an "irrational" object that satisfies the equation $x^2 = 2$.

Just as irrational numbers such as $\sqrt{2}$ lead to the construction of the real numbers, the imaginary number i leads to the construction of the complex numbers.

Definition 7.1.1

Complex Number, Standard Form, Equality of Complex Numbers

A **complex number** in **standard form** is an expression of the form $x + yi$ where $x, y \in \mathbb{R}$ and i satisfies $i^2 = -1$. The set of all complex numbers is denoted by

$$\mathbb{C} = \{x + yi \,|\, x, y \in \mathbb{R}\}.$$

Note that if we are given a complex number $z = x + yi$ in standard form, then we may safely assume that $x, y \in \mathbb{R}$.

Example 7.1.2

We have that $1 + 2i$, $4\pi + \sqrt{2}i$ and $3 - 2i$ are all in $\mathbb{C}$.

To adhere to Definition 7.1.1, we should write $3 + (-2)i$ in Example 7.1.2, but for convenience, we will write $x - yi$ instead of $x + (-y)i$, so we will consider $3 - 2i$ to be in standard form.

Example 7.1.3

If $x \in \mathbb{R}$ then we can express it as $x + 0i$ and in this way view every real number as a complex number. So, we now have

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

It should be apparent that a complex number has two "parts". This motivates the next definition.

Definition 7.1.4

Real Part, Imaginary Part, Purely Imaginary

Let $z = x + yi \in \mathbb{C}$ with $x, y \in \mathbb{R}$. We call x the **real part** of z and we call y the **imaginary part** of z:

$$\begin{aligned} x &= \mathrm{Re}(z) \quad \text{(sometimes written as } \Re(z)\text{)} \\ y &= \mathrm{Im}(z) \quad \text{(sometimes written as } \Im(z)\text{)}. \end{aligned}$$

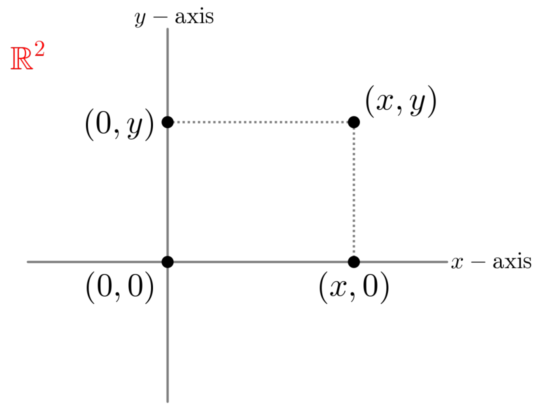
If $x = 0$, then we say z is **purely imaginary**. We also simply write $z = yi$ instead of $z = 0 + yi$.

Example 7.1.5

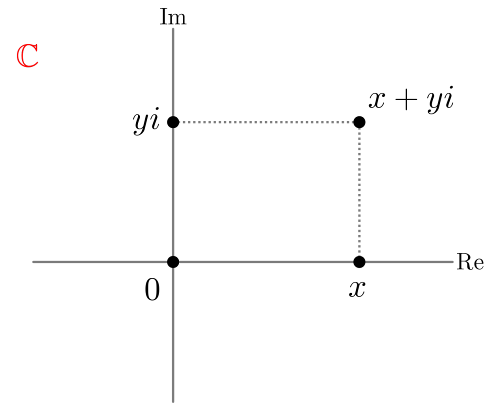
We have $\mathrm{Re}(3 - 4i) = 3$ and $\mathrm{Im}(3 - 4i) = -4$.

It is important to note that $\mathrm{Im}(3 - 4i) \neq -4i$. By definition, for any $z \in \mathbb{C}$ we have $\mathrm{Re}(z) \in \mathbb{R}$ and $\mathrm{Im}(z) \in \mathbb{R}$, that is, both the real and imaginary parts of a complex number are real numbers.

Geometrically, we interpret the set of real numbers as a line, called the real line. Given that $\mathbb{R} \subseteq \mathbb{C}$ and that there are complex numbers that are not real, the set of complex numbers should be "bigger" than a line. In fact, the set of complex numbers is a plane, much like the xy–plane[1] as shown in Figure 7.1.1. We "identify" the complex number $x + yi \in \mathbb{C}$ with the point $(x, y) \in \mathbb{R}^2$. In this sense, the complex plane is simply a "relabelling" of the xy–plane. The x–axis in the xy–plane corresponds to the real axis in the complex plane which contains the real numbers, and the y–axis of the xy–plane corresponds to the imaginary axis in the complex plane which contains the purely imaginary numbers. Note we will often label the real axis as "Re" and the imaginary axis as "Im".



(a) The xy–plane, also called $\mathbb{R}^2$.



(b) The complex plane $\mathbb{C}$.

Figure 7.1.1: The complex plane, $\mathbb{C}$, can be thought of as a relabelling of $\mathbb{R}^2$, where we rename the point (x, y) as the complex number $x + yi$.

Now we define the basic algebraic operations on complex numbers.

Definition 7.1.6
Equality

Two complex numbers $z = x + yi$ and $w = u + vi$ in standard form are **equal** if and only if $x = u$ and $y = v$, that is, if and only if $\mathrm{Re}(z) = \mathrm{Re}(w)$ and $\mathrm{Im}(z) = \mathrm{Im}(w)$.

Simply put, two complex numbers are equal if they have the same real parts and the same imaginary parts.

Definition 7.1.7
Addition, Subtraction, Multiplication

Let $z = x + yi$ and $w = u + vi$ be two complex numbers in standard form. We define **addition**, **subtraction** and **multiplication**, respectively, by

$$\begin{aligned} z + w &= (x + yi) + (u + vi) = (x + u) + (y + v)i \\ z - w &= (x + yi) - (u + vi) = (x - u) + (y - v)i \\ zw &= (x + yi)(u + vi) = (xu - yv) + (xv + yu)i. \end{aligned}$$

To add (resp. subtract) two complex numbers, we simply add (resp. subtract) the real parts and add the imaginary parts. With our definition of multiplication, we can verify

[1]To be consistent with our previous work, we should say the x_1x_2–plane, but since complex numbers only have two parts (a real part and an imaginary part), we will simply use x and y.

that $i^2 = -1$:

$$i^2 = (i)(i) = (0+1i)(0+1i) = (0(0)-1(1)) + (0(1)+1(0))i = -1+0i = -1.$$

There is no need to memorize the formula for multiplication of complex numbers. Using the fact that $i^2 = -1$, we can simply do a binomial expansion:

$$\begin{aligned}(x+yi)(u+vi) &= xu + xvi + yui + yvi^2 \\ &= xu + xvi + yui - yv \\ &= (xu - yv) + (xv + yu)i.\end{aligned}$$

We also see that

$$(-1)(u+vi) = (-1+0i)(u+vi) = -u - vi + 0ui + 0vi^2 = -u - vi,$$

from which it follows that $z - w = z + (-1)w$.

Example 7.1.8 Let $z = 3-2i$ and $w = -2+i$. Compute $z+w$, $z-w$ and zw. Express your answers in standard form.

Solution: We have

$$\begin{aligned}z+w &= (3-2i)+(-2+i) = (3+(-2)) + (-2+1)i = 1-i \\ z-w &= (3-2i)-(-2+i) = (3-(-2)) + (-2-1)i = 5-3i \\ zw &= (3-2i)(-2+i) = -6+3i+4i-2i^2 = -6+3i+4i+2 = -4+7i.\end{aligned}$$

Our geometric interpretation of addition is similar to that of vectors in $\mathbb{R}^2$.

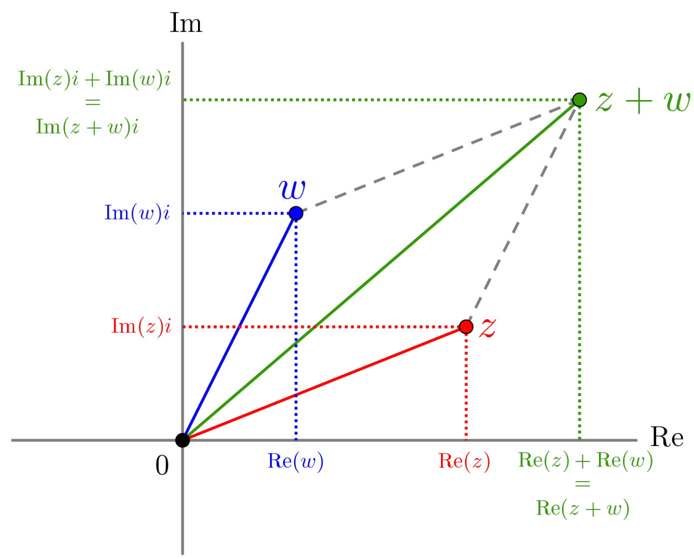


Figure 7.1.2: Visually interpreting complex addition.

Figure 7.1.2 shows that the complex numbers $0, z, w$ and $z+w$ determine a parallelogram with the line segment between 0 and $z+w$ as one of the diagonals. It is a good idea to compare Figure 7.1.2 with Figure 1.1.3.

Exercise 113 Show that our definition of addition and multiplication of complex numbers is consistent with the addition and multiplication of real numbers. That is, show that the sum and product of two real numbers x and y is the same as the sum and product of $x = x + 0i$ and $y = y + 0i$.

We now look at division of complex numbers. To start, we have to define $w = \frac{1}{z}$ for a nonzero $z = x + yi \in \mathbb{C}$. The key to doing this is the familiar trick of rationalizing the denominator. Pretending that i behaves like $\sqrt{-1}$, we can view $x - yi$ as being the *conjugate* of $x + yi$. Thus,

$$\begin{aligned}\frac{1}{z} &= \frac{1}{x+yi}\\ &= \frac{1}{x+yi} \cdot \frac{x-yi}{x-yi}\\ &= \frac{x-yi}{(x+yi)(x-yi)}\\ &= \frac{x-yi}{x^2 - xyi + xyi - y^2i^2}\\ &= \frac{x-yi}{x^2+y^2}\\ &= \frac{x}{x^2+y^2} - \frac{y}{x^2+y^2}i.\end{aligned}$$

The next exercise confirms that this expression works as one would expect of $w = \frac{1}{z}$.

Exercise 114 Let $z = x + yi$ and let $w = \dfrac{x}{x^2+y^2} - \dfrac{y}{x^2+y^2}i$. Show that $zw = 1$.

Note that when we multiplied the numerator and denominator by $x - yi$, the denominator turned into $(x + yi)(x - yi) = x^2 + y^2 \in \mathbb{R}$. This allowed us to put the quotient into standard form. We can now divide any complex number by any nonzero complex number by following this process. Here is the formal definition.

Definition 7.1.9
Division

Let $z = x + yi$ and $w = u + vi$ be two complex numbers in standard form. If $w \neq 0 + 0i$, We define **division** by

$$\frac{z}{w} = \frac{xu+yv}{u^2+v^2} + \frac{yu-xv}{u^2+v^2}i.$$

You should **not** memorize this definition. Instead, to compute z/w, simply multiply the numerator and denominator by the *conjugate* of w as illustrated in the next example.

Example 7.1.10 With $z = 3 - 2i$ and $w = -2 + i$, compute z/w in standard form.

Solution: We have

$$\frac{z}{w} = \frac{3-2i}{-2+i} = \frac{3-2i}{-2+i}\left(\frac{-2-i}{-2-i}\right) = \frac{-6-3i+4i+2i^2}{4+2i-2i-i^2} = \frac{-8+i}{4+1} = -\frac{8}{5} + \frac{1}{5}i.$$

Exercise 115

Express

$$\frac{(1-2i)-(3+4i)}{5-6i}$$

in standard form.

We finally discuss powers of a complex number.

Definition 7.1.11 **Integer Powers of Complex Numbers**

Let $z \in \mathbb{C}$. We define $z^1 = z$, and for any integer $k \geq 2$, $z^k = z^{k-1}z$. Provided $z \neq 0$, we additionally have $z^0 = 1$ and $z^{-k} = 1/z^k$ for any $k \geq 0$. In particular, $z^{-1} = 1/z$.

Notice that for integer powers of complex numbers, we have behaviour analogous to that of integer powers of real numbers. However, things become more complicated when the power of a complex number is not an integer, but rather any rational number, any real number, or even any complex number. Exploring such ideas is left to later courses.

The next theorem summarizes the rules of arithmetic in $\mathbb{C}$ and confirms that everything behaves as expected.

Theorem 7.1.12 **(Properties of Arithmetic in $\mathbb{C}$)**

Let $u, v, z \in \mathbb{C}$. Then

(a) $u + v = v + u$ — addition is commutative

(b) $(u + v) + z = u + (v + z)$ — addition is associative

(c) $z + 0 = z$ — 0 is the additive identity

(d) $z + (-z) = 0$ — $-z$ is the additive inverse of z

(e) $uv = vu$ — multiplication is commutative

(f) $(uv)z = u(vz)$ — multiplication is associative

(g) $z(1) = z$ — 1 is the multiplicative identity

(h) for $z \neq 0$, $z^{-1}z = 1$ — z^{-1} is the multiplicative inverse of $z \neq 0$

(i) $z(u + v) = zu + zv$ — distributive law

Section 7.1 Problems

7.1.1. For each of

$$\text{(a) } z = 3,\ w = 4i, \qquad \text{(b) } z = 2 + i,\ w = 3 - 2i,$$

evaluate the following.

(i) $\mathrm{Re}(z)$.

(ii) $\mathrm{Im}(w)$.

(iii) $z + w$.

(iv) $z - w$.

(v) zw.

(vi) $\dfrac{w}{z}$.

7.1.2. Write the following expressions in standard form.

(a) $\dfrac{(1-2i)+(2+3i)}{(5-6i)(-1+i)}$.

(b) $i\mathrm{Re}(4-6i) - \mathrm{Im}(2-3i)$.

7.1.3. Find all $z \in \mathbb{C}$ satisfying $z^2 = 21 + 20i$. [**Hint**: Let $z = a + bi$ with $a, b \in \mathbb{R}$.]

7.1.4. Let $\alpha \in \mathbb{R}$ and suppose $z \in \mathbb{C}$ satisfies the equation $(1 - \alpha i)z = \alpha - 9i$. Find all values of α so that $z \in \mathbb{R}$.

7.2 Conjugate and Modulus

In Section 7.1, we defined complex numbers and defined the operations of addition, subtraction, multiplication and division. To perform division, we saw that multiplying $x + yi$ by $x - yi$ was useful since it allowed us to write the quotient of two complex number in standard form. We now formally define the conjugate of a complex number.

Definition 7.2.1 **Complex Conjugate**

The **complex conjugate** of $z = x + yi$ with $x, y \in \mathbb{R}$ is $\overline{z} = x - yi$.

Note that we will often simply say conjugate rather than complex conjugate when it is clear that we mean complex conjugate.

Example 7.2.2

We have

- $\overline{1 + 3i} = 1 - 3i$
- $\overline{\sqrt{2}i} = -\sqrt{2}i$
- $\overline{-4} = -4$.

The conjugate enjoys some very natural properties as summarized in the following theorem.

Theorem 7.2.3

(Properties of Conjugates)

Let $z, w \in \mathbb{C}$ with $z = x + yi$ where $x, y \in \mathbb{R}$. Then

(a) $\overline{\overline{z}} = z$.

(b) $z + \overline{z} = 2x = 2\,\mathrm{Re}(z)$.

(c) $z - \overline{z} = 2yi = 2i\,\mathrm{Im}(z)$.

(d) $z \in \mathbb{R} \iff \overline{z} = z$.

(e) z is purely imaginary if and only if $\overline{z} = -z$.

(f) $\overline{z + w} = \overline{z} + \overline{w}$.

(g) $\overline{zw} = \overline{z}\,\overline{w}$.

(h) $\overline{\left(\dfrac{z}{w}\right)} = \dfrac{\overline{z}}{\overline{w}}$ provided $w \neq 0$.

(i) $z\overline{z} = x^2 + y^2$.

Proof: We prove (f) and leave the rest as an exercise. Let $z, w \in \mathbb{C}$ with $z = x + yi$ and $w = u + vi$ where $x, y, u, v \in \mathbb{R}$. Then

$$\overline{z + w} = \overline{(x + yi) + (u + vi)}$$

$$
\begin{aligned}
&= \overline{(x+u)+(y+v)i} \\
&= (x+u)-(y+v)i \\
&= (x-yi)+(u-vi) \\
&= \overline{z}+\overline{w}.
\end{aligned}
$$
□

Appealing to our geometric understanding, Figure 7.2.1 shows that we can view the conjugate as a reflection in the real axis. In particular, we see that if z is real, then $\overline{z} = z$ (Theorem 7.2.3(d)) and if z is purely imaginary, then $\overline{z} = -z$ (Theorem 7.2.3(e)).

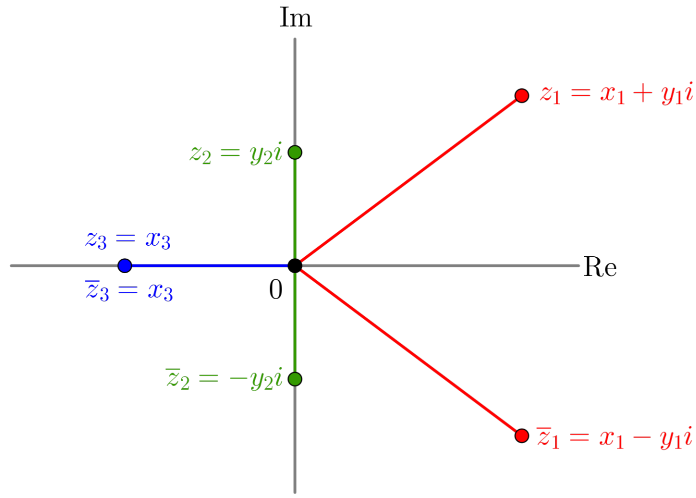


Figure 7.2.1: The conjugate of a complex number z is a refection of z in the real axis.

We note that (f) and (g) of Theorem 7.2.3 can be generalized to more than two complex numbers. For $z_1, \ldots, z_k \in \mathbb{C}$, we have

$$
\begin{aligned}
\overline{z_1 + \cdots + z_k} &= \overline{z}_1 + \cdots + \overline{z}_k \\
\overline{z_1 \cdots z_k} &= \overline{z}_1 \cdots \overline{z}_k.
\end{aligned}
$$

If $z_1 = \cdots = z_k = z$, then our second equation above gives

$$\overline{z^k} = \overline{z}^k$$

for any positive integer k. Additionally, for $z \neq 0$ and any integer $k \geq 0$, we use Theore 7.2.3(h) to obtain

$$\overline{z^{-k}} = \overline{\left(\frac{1}{z^k}\right)} = \frac{\overline{1}}{\overline{z^k}} = \frac{1}{\overline{z}^k} = (\overline{z})^{-k}.$$

Thus we have that

$$\overline{z^k} = \overline{z}^k$$

for any integer k, where we require $z \neq 0$ whenever $k \leq 0$.

Recall that the real numbers lie on a line (called the real line). Let $x, y \in \mathbb{R}$. If x is to the left of y on the real line, then we say $x < y$, and if x is to the right of y, then we say that $x > y$. If x is not to the right of y, then we say that $x \leq y$ and if x is not to the left of y, then we say that $x \geq y$. Thus, we can *order* the real numbers. However, we have come to understand that the complex numbers form a plane rather than a line, so we are not able to order the complex numbers as we do the real numbers. For example, we cannot say $1+i \leq 3i$ nor can we say $3i \leq 1+i$. However, the following definition will lead to a way for us to compare complex numbers.

Definition 7.2.4
Modulus

The **modulus** of $z = x+yi$ with $x, y \in \mathbb{R}$ is the nonnegative real number $|z| = \sqrt{x^2+y^2}$.

Example 7.2.5

We have

- $|1+i| = \sqrt{1^2+1^2} = \sqrt{2}$
- $|3i| = \sqrt{0^2+3^2} = \sqrt{9} = 3$
- $|-4| = \sqrt{(-4)^2+0^2} = \sqrt{16} = 4$

Let $x \in \mathbb{R}$ Then $x \in \mathbb{C}$ since $\mathbb{R} \subseteq \mathbb{C}$. Thus the modulus of x is given by

$$\underbrace{|x|}_{\text{modulus}} = |x+0i| = \sqrt{x^2+0^2} = \sqrt{x^2} = \underbrace{|x|}_{\text{absolute value}}.$$

We see that for $x \in \mathbb{R}$, the modulus of x is the absolute value of x. Thus the modulus is the extension of the absolute value to the complex numbers which is why we have chosen the same notation. We will thus interpret the modulus of $z \in \mathbb{C}$ to be the *size* or *magnitude* of z, just like the absolute value of $x \in \mathbb{R}$ can be interpreted as the size or magnitude of x.

As mentioned above, we cannot directly compare the complex numbers $1+i$ and $3i$ as we would with real numbers. However, the modulus does give us a way to indirectly compare these numbers: we have that

$$|1+i| = \sqrt{2} < 3 = |3i|.$$

As we are viewing the modulus of a complex number to be the extension of the absolute value of a real number, many of the properties listed in the following theorem should come as no surprise.

Theorem 7.2.6

(Properties of Modulus)

Let $z, w \in \mathbb{C}$. Then

(a) $|z| = 0 \iff z = 0$

(b) $|\overline{z}| = |z|$

(c) $z\overline{z} = |z|^2$

(d) $|zw| = |z||w|$

(e) $\left|\dfrac{z}{w}\right| = \dfrac{|z|}{|w|}$ provided $w \neq 0$

(f) $|z+w| \leq |z|+|w|$ which is known as the Triangle Inequality

Proof: We prove (d) and leave the rest as an exercise. Let $z, w \in \mathbb{C}$. We have

$$\begin{aligned} |zw|^2 &= (zw)\overline{zw} && \text{by (c)} \\ &= zw\overline{z}\,\overline{w} && \text{by Theorem 7.2.3(g)} \\ &= z\overline{z}w\overline{w} \\ &= |z|^2|w|^2 && \text{by (c)} \\ &= (|z||w|)^2 \end{aligned}$$

Thus $|zw|^2 = (|z||w|)^2$. Since the modulus of a complex number is never negative, we can take square roots of both sides to obtain $|zw| = |z||w|$. □

Note that for a complex number $z \neq 0$, Theorem 7.2.6(c) shows how the conjugate and the modulus combine to give us an efficient way to write z^{-1}:

$$z^{-1} = \frac{1}{z} = \frac{\overline{z}}{z\overline{z}} = \frac{\overline{z}}{|z|^2}.$$

Example 7.2.7 For $z = 2 - 5i$, we have that

$$z^{-1} = \frac{1}{2-5i} = \frac{\overline{2-5i}}{|2-5i|^2} = \frac{2+5i}{2^2+(-5)^2} = \frac{2}{29} + \frac{5}{29}i.$$

We note that Theorem 7.2.6(d) can be generalized to more than two complex numbers. For $z_1, \ldots, z_k \in \mathbb{C}$, we have

$$|z_1 \cdots z_k| = |z_1| \cdots |z_k|.$$

In particular, for $z_1 = \cdots = z_k = z$, we have that

$$|z^k| = |z|^k$$

for any positive integer k. In addition, for $z \neq 0$ and any integer $k \geq 0$, we can use Theorem 7.2.6(e) to obtain

$$|z^{-k}| = \left|\frac{1}{z^k}\right| = \frac{|1|}{|z^k|} = \frac{1}{|z|^k} = |z|^{-k}.$$

Thus we see that

$$|z^k| = |z|^k$$

for all integers k, with $k \leq 0$ requiring $z \neq 0$.

Figure 7.2.2 gives us a geometric understanding of the modulus. We see that $|z|$ is the distance between 0 and z, and that $|z| = |\overline{z}|$. We also notice that any $w \in \mathbb{C}$ lying inside the circle of radius $|z|$ centered about 0 will have modulus $|w| < |z|$, any $w \in \mathbb{C}$ lying on this circle will have satisfy $|w| = |z|$ and any $w \in \mathbb{C}$ lying outside the circle will be such that $|w| > |z|$.

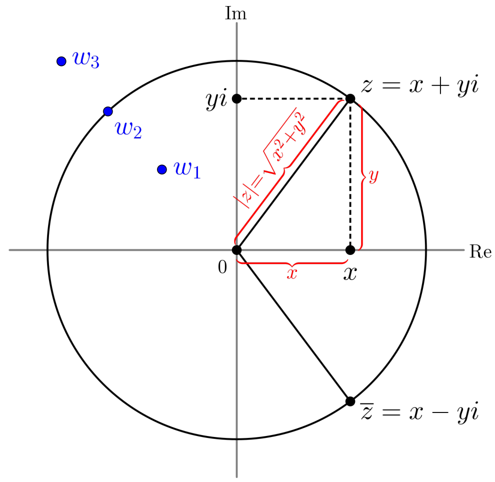


Figure 7.2.2: Visually interpreting the complex conjugate and the modulus of a complex number. Note that $|\overline{z}| = |z|$, $|w_1| < |z|$, $|w_2| = |z|$ and $|w_3| > |z|$.

We also observe in Figure 7.2.2 that for any $r \in \mathbb{R}$ with $r > 0$, there are infinitely many $z \in \mathbb{C}$ such that $|z| = r$. Compare this with the fact that there are only two $x \in \mathbb{R}$ with $|x| = r > 0$, namely $x = \pm r$. Indeed, a circle of radius $r > 0$ centred about 0 in the complex plane intersects the real axis in exactly two points: $z = r$ and $z = -r$.

Finally, we look at the triangle determined by complex numbers 0, z and $z + w$ as illustrated in Figure 7.2.3. Comparing Figure 7.2.3 to Figure 1.3.3 will help reinforce the many similarities between the complex plane $\mathbb{C}$ and $\mathbb{R}^2$.

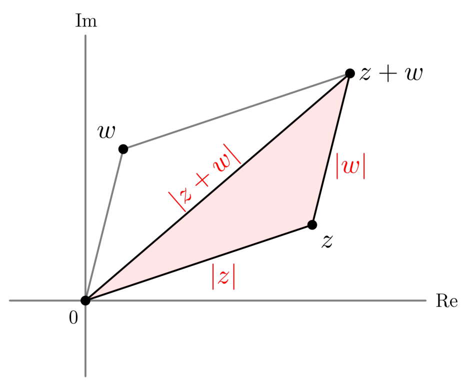


Figure 7.2.3: Visualizing the Triangle Inequality.

Since the length of any one side of a triangle cannot exceed the sum of the other two sides (or else the triangle wouldn't "close"), we observe from Figure 7.2.3 that

$$|z + w| \leq |z| + |w|.$$

Section 7.2 Problems

7.2.1. For each of

$$\text{(a) } z = 1 + i,\ w = 2 + 2i, \qquad \text{(b) } z = 2 + i,\ w = 3 - 2i,$$

evaluate the following.

(i) $\overline{z}$.

(ii) $|z|$.

(iii) $|w|$.

(iv) $|\overline{z}|$.

(v) $|z + w|$.

(vi) $|z| + |w|$.

(vii) $|zw|$.

(viii) $\left|\dfrac{z}{w}\right|$.

7.2.2. Express

$$|3 + 4i| \left(\overline{1 - 2i}\right) + (2 + 3|i|) \left(\overline{3i + 2}\right)$$

in standard form.

7.2.3. Find all $z \in \mathbb{C}$ satisfying $3z^2 = 4\overline{z}$.

7.2.4. (a) Find all $z \in \mathbb{C}$ such that $|z| = |\overline{z} + i|$.

(b) Find all $z \in \mathbb{C}$ such that $|z| \leq |z + 2|$.

7.2.5. Find all $z \in \mathbb{C}$ satisfying $\overline{z} + 2z = |z - 3|$.

7.2.6. Let $z, w \in \mathbb{C}$. Prove that $|z + w|^2 + |z - w|^2 = 2(|z|^2 + |w|^2)$.

7.3 Polar Form

We now look at another way that we can represent complex numbers that will help us gain a geometric understanding of complex multiplication. Consider a nonzero complex number $z = x + yi$ in standard form. Let $r = |z| > 0$ and let θ denote the angle the line segment from 0 to z makes with the positive real axis, measured counterclockwise. We refer to $r > 0$ as the *radius* of z and θ as an *argument* of z.

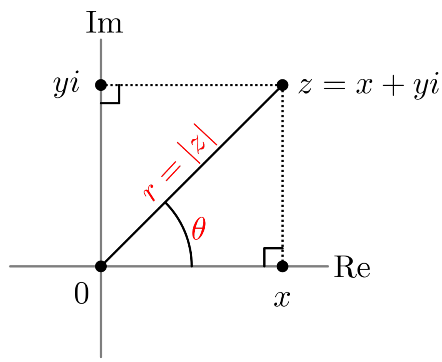


Figure 7.3.1: Depicting a complex number z, its modulus r and an argument θ.

Given $z = x + yi \neq 0$ in standard form, we compute $r = |z| = \sqrt{x^2 + y^2} > 0$ and we compute θ using

$$\cos\theta = \frac{x}{r} \quad \text{and} \quad \sin\theta = \frac{y}{r}.$$

It follows that

$$x = r\cos\theta \quad \text{and} \quad y = r\sin\theta$$

from which we obtain

$$z = x + yi = (r\cos\theta) + (r\sin\theta)i = r(\cos\theta + i\sin\theta).$$

Note that $|\cos\theta + i\sin\theta| = \sqrt{\cos^2\theta + \sin^2\theta} = 1$, and as a result, we may understand an argument of a complex number z as giving us a point on a circle of radius 1 to move towards (that is measured counterclockwise from the positive real axis), while $r > 0$ tell us how far to move in that direction to reach z. This is illustrated in Figure 7.3.2.

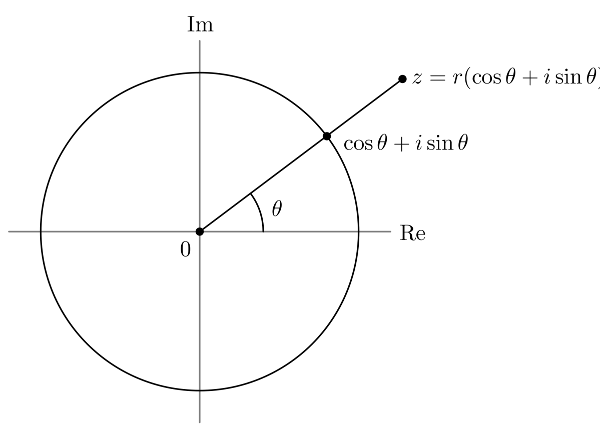


Figure 7.3.2: Using r and θ to locate a complex number. Here, $r > 1$.

So far, we have considered complex numbers $z \neq 0$. For $z = 0$, it is clear that $r = 0$ so that $0 = 0(\cos\theta + i\sin\theta)$ for any $\theta \in \mathbb{R}$.

Definition 7.3.1

Polar Form, Radius, Argument

The **polar form** of a complex number z is given by

$$z = r(\cos\theta + i\sin\theta)$$

where $r = |z|$ is the **radius** and θ is an **argument** of z.

We typically write $\cos\theta + i\sin\theta$ rather than $\cos\theta + (\sin\theta)i$ to avoid the extra brackets. For standard form, we still write $x + yi$ although it is not wrong to write $x + iy$. Note that unlike standard form, z does not have a unique polar form. Recall that for any $k \in \mathbb{Z}$,

$$\cos\theta = \cos(\theta + 2k\pi) \quad \text{and} \quad \sin\theta = \sin(\theta + 2k\pi)$$

so

$$r(\cos\theta + i\sin\theta) = r\big(\cos(\theta + 2k\pi) + i\sin(\theta + 2k\pi)\big)$$

for any $k \in \mathbb{Z}$.

Example 7.3.2

Write the following complex numbers in polar form.

(a) $1 + \sqrt{3}i$

(b) $7 + 7i$

Solution:

(a) We have $r = |1 + \sqrt{3}i| = \sqrt{1^2 + (\sqrt{3})^2} = \sqrt{1+3} = \sqrt{4} = 2$. Thus, factoring $r = 2$ out of $1 + \sqrt{3}i$ gives

$$1 + \sqrt{3}i = 2\left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right).$$

As this is of the form $r(\cos\theta + i\sin\theta)$, we have that $\cos\theta = \frac{1}{2}$ and $\sin\theta = \frac{\sqrt{3}}{2}$. We thus take $\theta = \frac{\pi}{3}$ so

$$1 + \sqrt{3}i = 2\left(\cos\frac{\pi}{3} + i\sin\frac{\pi}{3}\right).$$

(b) Since $r = |7 + 7i| = \sqrt{7^2 + 7^2} = \sqrt{2(49)} = 7\sqrt{2}$, we have that

$$7 + 7i = 7\sqrt{2}\left(\frac{7}{7\sqrt{2}} + \frac{7}{7\sqrt{2}}i\right) = 7\sqrt{2}\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i\right)$$

so $\cos\theta = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$ and $\sin\theta = \frac{\sqrt{2}}{2}$. Thus we take $\theta = \frac{\pi}{4}$ to obtain

$$7 + 7i = 7\sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right).$$

Note that we can add 2π to either of our above arguments to obtain

$$1+\sqrt{3}i = 2\left(\cos\frac{7\pi}{3} + i\sin\frac{7\pi}{3}\right)$$
$$7+7i = 7\sqrt{2}\left(\cos\frac{9\pi}{4} + i\sin\frac{9\pi}{4}\right)$$

which verifies that the polar form of a complex number is not unique. Normally, we choose our arguments θ such that $0 \le \theta < 2\pi$ or $-\pi < \theta \le \pi$ to avoid having these multiple representations.

We have seen that converting a complex number from standard form to polar form is a bit computational, however the next example shows it is quite easy to convert from polar form back to standard form.

Example 7.3.3 Write $3\big(\cos\left(\frac{5\pi}{6}\right) + i\sin\left(\frac{5\pi}{6}\right)\big)$ in standard form.

Solution: We have

$$3\left(\cos\left(\frac{5\pi}{6}\right) + i\sin\left(\frac{5\pi}{6}\right)\right) = 3\left(-\frac{\sqrt{3}}{2} + \frac{1}{2}i\right) = -\frac{3\sqrt{3}}{2} + \frac{3}{2}i.$$

The following theorem shows how easy it is to multiply complex numbers when they are in polar form.

Theorem 7.3.4 Let $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ be two complex numbers in polar form. Then

$$z_1 z_2 = r_1 r_2\big(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)\big).$$

Proof: Recall the angle sum formulas

$$\cos(\theta_1+\theta_2) = \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2$$
$$\sin(\theta_1+\theta_2) = \sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2$$

If

$$z_1 = r_1(\cos\theta_1 + i\sin\theta_1) \quad \text{and} \quad z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$$

are two complex numbers in polar form, then

$$\begin{aligned} z_1 z_2 &= \big(r_1(\cos\theta_1 + i\sin\theta_1)\big)\big(r_2(\cos\theta_2 + i\sin\theta_2)\big) \\ &= r_1 r_2(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= r_1 r_2\big((\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)\big) \\ &= r_1 r_2\big(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)\big). \end{aligned}$$ □

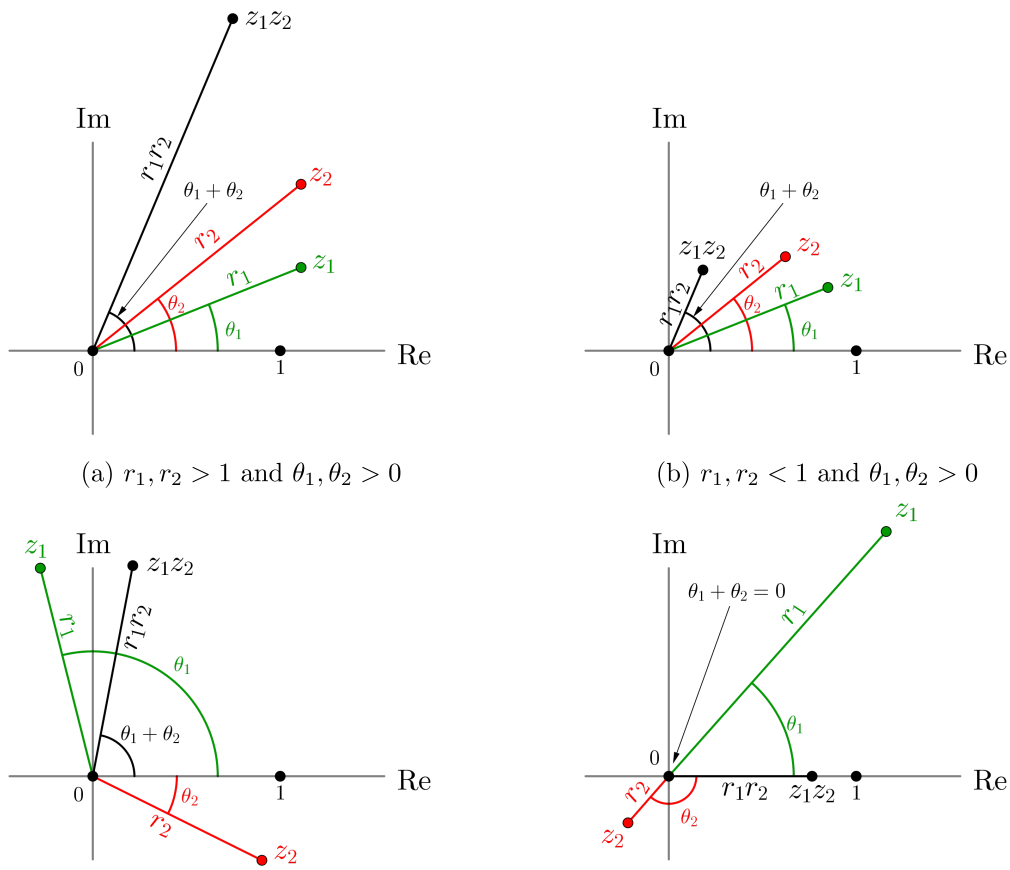
Example 7.3.5 Let $z_1 = 2\left(\cos\frac{\pi}{3} + i\sin\frac{\pi}{3}\right)$ and $z_2 = 7\sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$. Express $z_1 z_2$ in polar form.

Solution: We have

$$z_1 z_2 = 2(7\sqrt{2})\left(\cos\left(\frac{\pi}{3}+\frac{\pi}{4}\right)+i\sin\left(\frac{\pi}{3}+\frac{\pi}{4}\right)\right) = 14\sqrt{2}\left(\cos\frac{7\pi}{12}+i\sin\frac{7\pi}{12}\right).$$

Theorem 7.3.4 shows that when multiplying two complex numbers z_1 and z_2, both of which are in polar form, we simply multiply the moduli of z_1 and z_2 together to obtain the modulus of z_1z_2, and we simply add the given arguments of z_1 and z_2 together to derive an argument for z_1z_2. Although converting a complex number from standard form to polar form can be a bit tedious, the payoff is that we can avoid the binomial expansion needed to multiply two complex numbers in standard form. Instead we can compute the product of two moduli and the sum of two arguments, and both of these operations involve only real numbers.

Theorem 7.3.4 also leads to the geometric understanding of complex multiplication that we are looking for. We will view multiplication by a complex number $z = r(\cos\theta + i\sin\theta)$ as a counterclockwise rotation by an angle θ about 0, and a scaling by a factor of r. Note that a counterclockwise rotation by θ is a clockwise rotation by $-\theta$. Thus, if $\theta = -\frac{\pi}{4}$ for example, then multiplication by z can be viewed as a clockwise rotation by $\frac{\pi}{4}$ (plus a scaling by a factor of r). This is illustrated in Figure 7.3.3.



(a) $r_1, r_2 > 1$ and $\theta_1, \theta_2 > 0$

(b) $r_1, r_2 < 1$ and $\theta_1, \theta_2 > 0$

(c) $r_1 > 1, r_2 = 1$ and $\theta_1 > 0, \theta_2 < 0$

(d) $r_1 > 1, r_2 < 1$ and $\theta_1 = -\theta_2$

Figure 7.3.3: Multiplication $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ for various values of $r_1, r_2, \theta_1, \theta_2 \in \mathbb{R}$.

Exercise 116

Let $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ be two complex numbers in polar form with $z_2 \neq 0$ (from which it follows that $r_2 \neq 0$). Show that

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}\big(\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)\big).$$

7.3.1 Powers of Complex Numbers

Recall that if $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_1 + i\sin\theta_2)$ are two complex numbers in polar form, then by Theorem 7.3.4, we have that

$$z_1 z_2 = r_1 r_2\big(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\big).$$

Note that Theorem 7.3.4 generalizes to more than two complex numbers. If

$$z_1 = r_1(\cos\theta_1 + i\sin\theta_1), \dots, z_n = r_n(\cos\theta_n + i\sin\theta_n)$$

are n complex numbers in polar form, then repeated applications of Theorem 7.3.4 gives

$$z_1 \cdots z_n = r_1 \cdots r_n\big(\cos(\theta_1 + \cdots + \theta_n) + i\sin(\theta_1 + \cdots + \theta_n)\big). \tag{7.6}$$

Taking $z_1 = \cdots = z_n$ with their common value being $z = r(\cos\theta + i\sin\theta)$, (7.6) reduces to

$$z^n = r^n\big(\cos(n\theta) + i\sin(n\theta)\big). \tag{7.7}$$

Thus, for any $z \in \mathbb{C}$ and any $n \in \mathbb{N}$, (7.7) gives us a *very* fast way to compute z^n given that we have polar form of z.

Exercise 117

For $z = r(\cos\theta + i\sin\theta) \neq 0$, show that

$$z^{-1} = \frac{1}{z} = \frac{1}{r}\big(\cos(-\theta) + i\sin(-\theta)\big).$$

It follows from Exercise 117 that for a complex number $z \neq 0$, (7.7) holds for any $n \in \mathbb{Z}$. This gives the following important result.

Theorem 7.3.6 **(de Moivre's Theorem)**

If $z = r(\cos\theta + i\sin\theta) \neq 0$, then

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

for any $n \in \mathbb{Z}$.

Since de Moivre's Theorem is stated for $n \in \mathbb{Z}$, we have to allow for $n \leq 0$ and thus the restriction that $z \neq 0$. It is easy to verify that de Moivre's Theorem holds for $z = 0$ provided $n \geq 1$ since $z^n = 0$ in this case.

Example 7.3.7

Compute $(2+2i)^7$ using de Moivre's Theorem and express your answer in standard form.

Solution: We have $r = |2+2i| = \sqrt{4+4} = \sqrt{2(4)} = 2\sqrt{2}$ and so

$$2+2i = 2\sqrt{2}\left(\frac{2}{2\sqrt{2}} + \frac{2}{2\sqrt{2}}\,i\right) = 2\sqrt{2}\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\,i\right)$$

from which we find $\theta = \frac{\pi}{4}$. Thus

$$2+2i = 2\sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right).$$

Then

$$\begin{aligned}
(2+2i)^7 &= \left(2\sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)\right)^7 \\
&= (2\sqrt{2})^7\left(\cos\frac{7\pi}{4} + i\sin\frac{7\pi}{4}\right) \quad \text{by de Moivre's Theorem} \\
&= 1024\sqrt{2}\left(\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}\,i\right) \\
&= 1024 - 1024i.
\end{aligned}$$

Exercise 118

Compute $\left(\frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right)^{602}$ and express your answer in standard form.

It is hopefully apparent that trigonometry is playing a vital role here, so we include the complex version of the unit circle in Figure 7.3.4.

7.3.2 Complex Exponential Form

In this section, we introduce the notation $e^{i\theta}$ and briefly look at how it relates to polar form.

Definition 7.3.8
Complex Exponential Form

Let $\theta \in \mathbb{R}$. The expression $e^{i\theta}$ is defined to mean

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

If $z = r(\cos\theta + i\sin\theta)$ is the polar form of $z \in \mathbb{C}$, then $z = re^{i\theta}$ is the **complex exponential form** of z.

In MATH 115, the expression $e^{i\theta}$ will only be used as a short-hand for the complex number $\cos\theta + i\sin\theta$. However, you should know that it is possible to define a complex version of the exponential, sine and cosine functions. That is, we can make sense of e^z, $\sin z$ and $\cos z$ when $z \in \mathbb{C}$. If we do this carefully, we can then *obtain* the result that the exponential $e^{i\theta}$

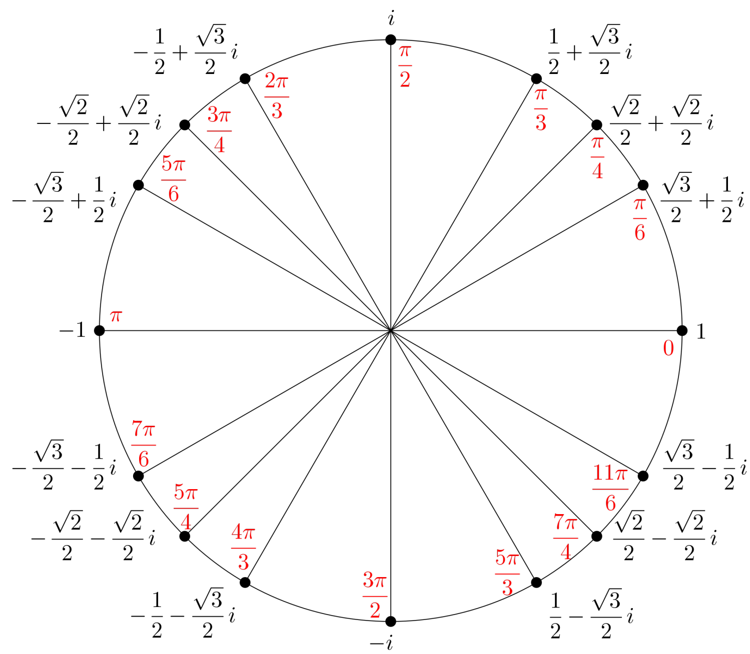


Figure 7.3.4: The unit circle in the complex plane.

is equal to $\cos\theta + i\sin\theta$ (rather than simply making it a definition as we do here). This surprising result, which equates an exponential value with a combination of trigonometric values, is known as *Euler's Formula*.

Example 7.3.9

Since

$$e^{i\frac{\pi}{4}} = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4} = \frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}},$$

we see that the complex exponential form of $z = \frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}$ is $z = e^{i\frac{\pi}{4}}$.

Similarly, for every $k \in \mathbb{Z}$, we have

$$e^{i\left(\frac{\pi}{4}+2k\pi\right)} = \cos\left(\frac{\pi}{4}+2k\pi\right) + i\sin\left(\frac{\pi}{4}+2k\pi\right) = \frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}} = e^{i\frac{\pi}{4}}.$$

Generalizing the previous example, we see that if $z = re^{i\theta}$ is a complex exponential form of z, then so is $z = re^{i(\theta+2k\pi)}$ for every $k \in \mathbb{Z}$. Thus, just like polar form, complex exponential form is not unique.

Example 7.3.10

(Euler's Identity)

We have

$$e^{i\pi} = \cos\pi + i\sin\pi = -1 + i(0) = -1.$$

So the complex exponential form of $z = -1$ is $z = e^{i\pi}$. That is, $-1 = e^{i\pi}$. This result is often re-written in the more suggestive form

$$e^{i\pi} + 1 = 0.$$

This equation, called *Euler's Identity*, is often regarded as one of the most beautiful equations in mathematics, since it relates the five important constants e, i, π, 1 and 0.

Exercise 119 Show that complex exponential forms of $z = 1$ and $w = i$ are given by $z = e^{i0}$ and $w = e^{i\frac{\pi}{2}}$, respectively.

As with the complex polar form $z = r(\cos\theta + i\sin\theta)$, the complex exponential form $z = re^{i\theta}$ allows us to perform complex multiplication very quickly, as the next theorem shows. The advantage to the complex exponential form is that it is more compact than the complex polar form. The next theorem also shows that $e^{i\theta}$ obeys the multiplication law of exponential functions, which justifies our choice of writing it as an exponential.

Theorem 7.3.11 Let $z_1 = r_1e^{i\theta_1}$ and $z_2 = r_2e^{i\theta_2}$ be complex exponential forms of $z_1, z_2 \in \mathbb{C}$. Then

$$z_1z_2 = r_1r_2e^{i(\theta_1+\theta_2)}.$$

Exercise 120 Provide a proof for Theorem 7.3.11.

The result of Theorem 7.3.11 generalizes to any n complex numbers where $n \in \mathbb{N}$. If

$$z_1 = r_1e^{i\theta_1}, \ldots, z_n = r_ne^{i\theta_n}$$

are the complex exponential forms of $z_1, \ldots, z_n \in \mathbb{C}$, then

$$z_1 \cdots z_n = r_1 \cdots r_n e^{i(\theta_1+\cdots+\theta_n)}. \tag{7.8}$$

In particular, if $z_1 = \cdots = z_n$ with their common value being $z = re^{i\theta}$, then (7.8) simplifies to

$$z^n = (re^{i\theta})^n = r^ne^{i(n\theta)}. \tag{7.9}$$

Note that we can also obtain (7.9) using de Moivre's Theorem. For $z = re^{i\theta} \in \mathbb{C}$,

$$z^n = (re^{i\theta})^n = (r(\cos\theta + i\sin\theta))^n = r^n(\cos\theta + i\sin\theta)^n = r^n(\cos n\theta + i\sin n\theta) = r^ne^{i(n\theta)}.$$

This identity is valid for all $n \in \mathbb{Z}$, provided that $z \neq 0$ whenever $n \leq 0$.

Section 7.3 Problems

7.3.1. Express the following complex numbers in (i) polar form and (ii) exponential form.

(a) $z = -i$.

(b) $z = -5$.

(c) $z = 3 + 4i$.

(d) $z = -3 + 4i$.

Note: In (c) and (d), give an approximate value for θ to 3 decimal points (in radians).

7.3.2. Express the following complex numbers in standard form.

(a) $z = 3e^{\frac{-5\pi}{6}i}$.

(b) $z = 5(\cos(\frac{5\pi}{2}) + i\sin(\frac{5\pi}{2}))$.

(c) $z = 3(\cos(\frac{5\pi}{3}) + i\sin(\frac{5\pi}{3})) \cdot 4(\cos(\frac{\pi}{6}) + i\sin(\frac{\pi}{6}))$.

(d) $z = \dfrac{\cos(\frac{7\pi}{4}) + i\sin(\frac{7\pi}{4})}{2\big(\cos(\frac{3\pi}{4}) + i\sin(\frac{3\pi}{4})\big)}$.

7.3.3. Let $z = \dfrac{\sqrt{2}}{2} - \dfrac{\sqrt{6}}{2}i$. Compute z^{10} in standard form using de Moivre's Theorem.

7.3.4. Use de Moivre's Theorem to show that

$$\cos(4\theta) = 8\cos^4\theta - 8\cos^2\theta + 1.$$

[**Hint:** Examine $(\cos\theta + i\sin\theta)^4$.]

7.4 Complex Polynomials

Definition 7.4.1

Real Polynomial, Variable, Coefficient, Root

We define a **real polynomial** of **degree** n by

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ with $a_n \neq 0$. We call x the **variable** and $a_0, a_1, \ldots, a_n$ the **coefficients**. A number c is a **root** (or a **zero**) of $p(x)$ if $p(c) = 0$.

You studied real polynomials in high school. Here we will also consider *complex polynomials*.

Definition 7.4.2

Complex Polynomial, Variable, Coefficient, Root

We define a **complex polynomial** of **degree** n by

$$p(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{C}$ with $a_n \neq 0$. We call z the **variable** and $a_0, a_1, \ldots, a_n$ the **coefficients**. A number c is a **root** (or a **zero**) of $p(z)$ if $p(c) = 0$.

Although the variable x is traditionally used for real polynomials while the variable z is normally used for complex polynomials, this is more of a convention and certainly not a requirement.

Example 7.4.3

The polynomial $p(z) = iz^3 - (1-i)z^2 + 3z + (4-i)$ is a complex polynomial of degree 3 with coefficients

$$a_3 = i, \quad a_2 = -(1-i), \quad a_1 = 3 \quad \text{and} \quad a_0 = 4 - i.$$

Since $\mathbb{R} \subseteq \mathbb{C}$, every real polynomial is in fact a complex polynomial. For example, $p(x) = x^2$ is a real polynomial, and thus a complex polynomial (the use of the variable x indicates that we are thinking of $p(x)$ first and foremost as a real polynomial). However, not every complex polynomial is a real polynomial: $p(z)$ as defined in Example 7.4.3 is not a real polynomial.

We define the basic operations of complex polynomials. These results also hold for real polynomials, where they should be familiar from high school. We begin with equality.

Definition 7.4.4

Equality

Let $p(z)$ and $q(z)$ be two complex polynomials of degree n with

$$p(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$
$$q(z) = b_n z^n + b_{n-1}z^{n-1} + \cdots + b_1 z + b_0$$

for some $a_0, \ldots, a_n, b_0, \ldots, b_n \in \mathbb{C}$. We say that $p(z)$ and $q(z)$ are **equal** if $a_k = b_k$ for $k = 0, 1, \ldots, n$.

We now turn to the standard operations of addition, subtraction and scalar multiplication.

Definition 7.4.5

Addition, Subtraction, Scalar Multiplication

Let $p(z)$ and $q(z)$ be two complex polynomials with

$$\begin{aligned}p(z) &= a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0\\ q(z) &= b_n z^n + b_{n-1} z^{n-1} + \cdots + b_1 z + b_0\end{aligned}$$

for some $a_0, \ldots, a_n, b_0, \ldots, b_n \in \mathbb{C}$. We define **addition** by

$$p(z) + q(z) = (a_n + b_n)z^n + (a_{n-1} + b_{n-1})z^{n-1} + \cdots + (a_1 + b_1)z + (a_0 + b_0),$$

we define **subtraction** by

$$p(z) - q(z) = (a_n - b_n)z^n + (a_{n-1} - b_{n-1})z^{n-1} + \cdots + (a_1 - b_1)z + (a_0 - b_0),$$

and for $k \in \mathbb{C}$, we define **scalar multiplication** by

$$kp(z) = ka_n z^n + ka_{n-1} z^{n-1} + \cdots + ka_1 z + ka_0.$$

Definition 7.4.5 makes no mention of the degree of $p(z)$ and $q(z)$, that is, we don't assume that $a_n \neq 0$ or that $b_n \neq 0$. This is because we can add polynomials of different degrees. For example, if $p(z) = z^2 + i$ and $q(z) = iz^3$, then we can write these as $p(z) = 0z^3 + 1z^2 + i$ and $q(z) = iz^3 + 0z^2 + 0$ to get that $p(z) + q(z) = iz^3 + z^2 + i$.

In words, Definition 7.4.5 says that we add (respectively subtract) polynomials by adding (respectively subtracting) their corresponding coefficients and that we multiply a polynomial by a complex number k by multiplying each coefficient of the polynomial by k.

As we have seen with vectors, matrices and linear transformations before, we have that

$$p(z) - q(z) = p(z) + (-1)q(z)).$$

Example 7.4.6

Let $p(z) = 3iz^2 + 4z - (1 + i)$ and $q(z) = 2z^2 + (2 - i)z + 5 + 2i$. Compute

(a) $p(z) + q(z)$.

(b) $(1 + i)p(z)$.

Solution:

(a) Adding corresponding coefficients, we have

$$\begin{aligned}p(z) + q(z) &= (3i + 2)z^2 + \big(4 + (2 - i)\big)z + \big(-(1 + i) + 2i\big)\\ &= (2 + 3i)z^2 + (6 - i)z + 4 + i.\end{aligned}$$

(b) Multiplying the coefficients of $p(z)$ by $1 + i$ gives

$$\begin{aligned}(1 + i)p(z) &= (1 + i)3iz^2 + (1 + i)4z - (1 + i)(1 + i)\\ &= (-3 + 3i)z^2 + (4 + 4i)z - 2i.\end{aligned}$$

Exercise 121

Let $p(z) = (1+i)z^4 - (2-i)z^2 + 4iz + 4$ and $q(z) = 5z^3 + (2+i)z^2 - 2 - i$. Compute

(a) $p(z) + q(z)$.

(b) $iq(z)$.

A fact learned in high school is that a real polynomial need not have a real root: consider $p(x) = x^2 + 1$ as an example - plotting the polynomial in the plane reveals a parabola that never touches the x-axis. However $p(x)$ is also a complex polynomial with two complex roots: $x = i$ and $x = -i$ since $p(\pm i) = (\pm i)^2 + 1 = -1 + 1 = 0$. The Fundamental Theorem of Algebra states that every non-constant complex polynomial will have at least one complex root. The proof of this Theorem requires a bit more knowledge of polynomials and is thus omitted.

Theorem 7.4.7 **(Fundamental Theorem of Algebra)**

Let $p(z)$ be a complex polynomial of degree at least 1 (that is, $p(z)$ is not a constant polynomial). Then $p(z)$ has at least 1 complex root.

Another (much more basic) theorem from algebra says that if a polynomial $p(z)$ of degree n has a root $c_1 \in \mathbb{C}$, then $p(z)$ can be factored as

$$p(z) = (z - c_1)q(z)$$

where $q(z)$ is a polynomial whose degree is $n-1$. If $q(z)$ is not a constant polynomial, then we can apply the Fundamental Theorem of Algebra again to conclude that $q(z)$ has a root (call it $c_2 \in \mathbb{C}$), and hence can itself be factored as

$$q(z) = (z - c_2)r(z),$$

where $r(z)$ is a polynomial of degree $n-2$. Hence

$$p(z) = (z - c_1)(z - c_2)r(z).$$

Continuing in this way, we eventually arrive at a factorization of $p(z)$ of the form

$$p(z) = (z - c_1)(z - c_2)\cdots(z - c_n)k$$

where $k \in \mathbb{C}$ is a constant.

This provides a slight strengthening of the Fundamental Theorem of Algebra and shows that a complex polynomial of degree $n \geq 1$ has n complex roots. However, these n roots need not be distinct. For example, the degree 6 polynomial $p(z) = (z-i)^2(z-1)^3(z-(2+i))$ has three distinct roots, i, 1 and $2+i$. The root i appears twice, and we say it has *multiplicity* equal to 2. Similarly, the root 1 appears three times, and we say it has *multiplicity* equal to 3. To summarize, we have the following result.

Theorem 7.4.8 Let $p(z)$ be a complex polynomial of degree $n \geq 1$. Then $p(z)$ has exactly n complex roots, if we count roots according to their multiplicities.

Example 7.4.9 The degree 10 polynomial $p(z) = 13(z-(4-i))^7(z+3i)^3$ has ten roots if we count multiplicities:

- the root $4-i$ has multiplicity 7, and
- the root $-3i$ has multiplicity 3.

We noted above that the real polynomial $p(x) = x^2+1$ has complex roots $\pm i$. That these two roots are complex conjugates of one another is not a coincidence.

Theorem 7.4.10 **(Conjugate Root Theorem)**

Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a real polynomial. If $w \in \mathbb{C}$ is a root of $p(x)$, then so too is $\overline{w}$.

Proof: Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a real polynomial and suppose $w \in \mathbb{C}$ is a root of $p(x)$. Then $p(w) = 0$, that is

$$a_n w^n + a_{n-1} w^{n-1} + \cdots + a_1 w + a_0 = 0.$$

Taking complex conjugates of both sides and using the fact that $0, a_0, a_1, \ldots, a_n \in \mathbb{R}$, we have

$$\begin{aligned}
\overline{a_n\ w^n + a_{n-1}\ w^{n-1} + \cdots + a_1\ w + a_0} &= \overline{0}\\
\overline{a_n\ w^n} + \overline{a_{n-1}\ w^{n-1}} + \cdots + \overline{a_1\ w} + \overline{a_0} &= \overline{0}\\
\overline{a_n}\ \overline{w^n} + \overline{a_{n-1}}\ \overline{w^{n-1}} + \cdots + \overline{a_1}\ \overline{w} + \overline{a_0} &= \overline{0}\\
\overline{a_n}\ \overline{w}^n + \overline{a_{n-1}}\ \overline{w}^{n-1} + \cdots + \overline{a_1}\ \overline{w} + \overline{a_0} &= \overline{0}\\
a_n\ \overline{w}^n + a_{n-1}\ \overline{w}^{n-1} + \cdots + a_1\ \overline{w} + a_0 &= 0.
\end{aligned}$$

Thus $p(\overline{w}) = 0$ so $\overline{w}$ is a root of $p(x)$. □

Note that we require $p(x)$ to be a real polynomial for Theorem 7.4.10 to hold. The complex polynomial

$$p(z) = z^2 + (2+3i)z - (5-i)$$

has roots $1-i$ and $-3-2i$, neither of which is a complex conjugate of the other.

Our next two examples deal with factoring polynomials. Before we begin, we briefly talk about square roots. As mentioned in Section 7.3, we can define e^z, $\sin z$ and $\cos z$ for complex numbers z. We can also do this for $\sqrt{z}$. Although we do not pursue this in any depth here, we will note that

$$\sqrt{-1} = i,$$

with the understanding that we are "sweeping a lot of details under the rug". Now, let $x, y \in \mathbb{R}$. Recall that for $x, y \geq 0$, we have that

$$\sqrt{xy} = \sqrt{x}\sqrt{y}. \tag{7.10}$$

It turns out that (7.10) holds if we allow *one* of x and y to be negative! Thus for $x \geq 0$, we have

$$\sqrt{-x} = \sqrt{x(-1)} = \sqrt{x}\sqrt{-1} = \sqrt{x}i.$$

So we can now evaluate the square root of any negative real number. For example, $\sqrt{-4} = 2i$ and $\sqrt{-7} = \sqrt{7}i$.

Note that we cannot apply (7.10) when *both* of x and y are negative, as evidenced by the following famous "proof" that $1 = -1$:

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = i(i) = i^2 = -1.$$

Why (7.10) doesn't hold when both $x, y < 0$ and how one extends the square root function to non-real numbers are topics explored in a complex analysis course.

Example 7.4.11

Let $p(x) = x^3 + 16x$. If $p(x) = 0$, then $0 = x^3 + 16x = x(x^2 + 16)$. Thus $x = 0$ or $x^2 + 16 = 0$. For $x^2 + 16 = 0$, we can use the quadratic formula:

$$\begin{aligned} x &= \frac{-0 \pm \sqrt{0^2 - 4(1)(16)}}{2(1)} \\ &= \pm\frac{\sqrt{-64}}{2} \\ &= \pm\frac{8i}{2} \\ &= \pm 4i. \end{aligned}$$

Thus the roots of $p(x)$ are 0, $4i$ and $-4i$. Note that given any of these roots, the complex conjugate of that root is also a root of $p(x)$.

Example 7.4.12

Let $z \in \mathbb{C}$ and consider the polynomial $p(z) = 3z^3 - az^2 - bz + 6b$ where $a, b \in \mathbb{R}$. It is known that $2 + 2i$ is a root of $p(z)$. Find a and b as well as the other roots of $p(z)$. Remember that if w is a root of $p(z)$, then $z - w$ is a factor of $p(z)$.

Solution: Since $p(z)$ has real coefficients and $2 + 2i$ is a root of $p(z)$, we have that $2 - 2i$ is also a root of $p(z)$ by Theorem 7.4.10. Since $p(z)$ has degree 3, there is a third root of $p(z)$ by Theorem 7.4.8. Let $w \in \mathbb{C}$ be this third root. Then

$$\begin{aligned} 3z^3 - az^2 - bz + 6b &= 3\big(z - (2+2i)\big)\big(z - (2-2i)\big)(z - w) \\ &= 3\big(z^2 - (2-2i)z - (2+2i)z + 8\big)(z - w) \\ &= 3\big(z^2 - 4z + 8\big)(z - w) \\ &= 3(z^3 - wz^2 - 4z^2 + 4wz + 8z - 8w) \\ &= 3\big(z^3 - (w+4)z^2 + (4w+8)z - 8w\big) \\ &= 3z^3 - (3w+12)z^2 + (12w+24)z - 24w \end{aligned}$$

Equating the z coefficients and the constant terms, we see that

$$-b = 12w + 24 \tag{7.11}$$

$$6b = -24w \tag{7.12}$$

From (7.12), we see that $b = -4w$ and substituting this into (7.11) gives $4w = 12w + 24$. Simplifying gives $8w = -24$, so $w = -3$. From $b = -4w$, we now have $b = 12$. Finally, equating the z^2 coefficients gives

$$a = 3w + 12 = 3(-3) + 12 = -9 + 12 = 3.$$

Thus $a = 3, b = 12$ and the other two roots are $2 - 2i$ and -3.

Section 7.4 Problems

7.4.1. Use the quadratic formula to find the complex roots of the following polynomials. Express your answer in standard form.

(a) $p(z) = z^2 + 3$.

(b) $p(z) = z^2 + z + 1$.

(c) $p(z) = 2z^2 - 3z + 4$.

7.4.2. Let $\alpha, \beta \in \mathbb{R}$ and consider the polynomial $p(z) = 2z^3 + \alpha z^2 + \beta z + 16 = 0$. Given that $p(z)$ has three roots and that one of them is $2 + 2\sqrt{3}i$, find:

(a) The other two roots of $p(z)$.

(b) α and β.

7.4.3. Can a complex polynomial of degree two have exactly one real root (and therefore exactly one non-real root)? Either give an example of such a polynomial or explain why no such polynomial can exist.

7.5 Complex nth Roots

In the previous section we saw that a complex polynomial of degree $n \geq 1$ has n complex roots, counted with multiplicities. However, given an arbitrary polynomial, it is not an easy task to actually find these n roots. In fact, more often than not, you will have to rely on numerical techniques to do so.

There are, fortunately, some exceptions. In this section we will learn how to find the roots of polynomials of the form $p(z) = z^n - w$, where $w \in \mathbb{C}$ is constant. Since any such root z_0 must satisfy $z_0^n = w$, we see that we're essentially asking for nth roots of w. Let's look at some examples.

Example 7.5.1 Find all $z \in \mathbb{C}$ satisfying the given equations.

(a) $z^2 = -2$.

(b) $z^2 = -7 + 24i$.

Notice that we are asking for the roots of the polynomials $z^2 + 2$ and $z^2 - (-7 + 24i)$, respectively.

Solution:

(a) The solutions are $z = \pm\sqrt{-2}$. By what we've discussed in the previous section, these are the two complex numbers $\sqrt{2}i$ and $-\sqrt{2}i$.

(b) Similarly, the solutions are $z = \pm\sqrt{-7 + 24i}$. However, it is not obvious how to express these complex numbers in standard form.

Here is one approach. Let $z = a + bi$ with $a, b \in \mathbb{R}$. Then the given equation becomes

$$z^2 = (a + bi)^2 = a^2 - b^2 + 2abi = -7 + 24i.$$

Equating real and imaginary parts gives

$$a^2 - b^2 = -7 \tag{7.13}$$
$$2ab = 24 \tag{7.14}$$

From (7.14), we have that $a, b \neq 0$, so $b = \frac{24}{2a} = \frac{12}{a}$. Substituting $b = \frac{12}{a}$ into (7.13) gives

$$\begin{aligned} a^2 - \left(\frac{12}{a}\right)^2 &= -7 \\ a^2 - \frac{144}{a^2} &= -7 \\ a^4 + 7a^2 - 144 &= 0 \\ (a^2 + 16)(a^2 - 9) &= 0 \\ (a^2 + 16)(a + 3)(a - 3) &= 0. \end{aligned}$$

Since $a \in \mathbb{R}$, $a^2 + 16 > 0$, so we conclude that $a + 3 = 0$ or $a - 3 = 0$ which gives $a = 3$ or $a = -3$. Since $b = \frac{12}{a}$, $b = \frac{12}{3} = 4$ or $b = \frac{12}{-3} = -4$. Thus $z = 3 + 4i$ or $z = -3 - 4i$.

The method illustrated in the previous example works decently well if the degree is $n = 2$, but for larger n, it quickly becomes impractical. Instead, we can use complex exponential form and de Moivre's theorem.

Recall that our problem is the following. We want to solve the equation

$$z^n = w$$

for $z \in \mathbb{C}$, assuming $w \in \mathbb{C}$ is given. Let $z = re^{i\theta}$ and $w = Re^{i\phi}$. From $z^n = w$, we find

$$r^n e^{i(n\theta)} = Re^{i\phi}.$$

Converting to polar form, we have

$$r^n(\cos(n\theta) + i\sin(n\theta)) = R(\cos\phi + i\sin(\phi)).$$

By equating the radii and the arguments, we obtain

$$r^n = R \quad \text{and} \quad n\theta = \phi + 2k\pi \text{ for some } k \in \mathbb{Z}.$$

Solving for r and θ, we get

$$r = R^{1/n} \quad \text{and} \quad \theta = \frac{\phi + 2k\pi}{n} \text{ for some } k \in \mathbb{Z}.$$

Note that since r and R are nonnegative real numbers, here $R^{1/n}$ is nonnegative real nth root evaluated in the usual way.

Since we are allowing k to be an arbitrary integer, there are infinitely many possible values for θ. It is tempting to think that there be infinitely many solutions to $z^n = w$ as a result, but in fact we only obtain finitely many solutions.

Theorem 7.5.2 Let $w = Re^{i\phi}$ be a nonzero complex number, and let n be a positive integer. There are precisely n distinct nth roots of w, and they are given by

$$z_k = R^{1/n} e^{i\frac{\phi+2k\pi}{n}}$$

for $k = 0, 1, \ldots, n-1$.

A few examples will show why we only need to consider $k = 0, 1, \ldots, n-1$.

Example 7.5.3 Find the third roots of 1, that is, find all $z \in \mathbb{C}$ such that $z^3 = 1$.

Solution: The equation we must solve is of the form $z^n = w$ with $w = 1$ and $n = 3$. In exponential form, $1 = 1(\cos 0 + i\sin 0) = 1e^{i0}$ so the 3rd roots of 1 are given by

$$z_k = 1^{1/3} e^{i\frac{0+2k\pi}{3}} = e^{i\frac{2k\pi}{3}}, \qquad \text{for } k = 0, 1, 2.$$

Thus,

$$z_0 = e^{i0} = \cos 0 + i\sin 0 = 1$$

$$z_1 = e^{i\frac{2\pi}{3}} = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}\,i$$
$$z_2 = e^{i\frac{4\pi}{3}} = \cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3} = -\frac{1}{2} - \frac{\sqrt{3}}{2}\,i.$$

The 3rd roots of 1 are therefore given by 1, $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$ and $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$. This means that

$$1^3 = \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right)^3 = \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}\,i\right)^3 = 1.$$

If we try to compute z_k for $k = -1$ or $k = 3$, we find

$$z_{-1} = e^{i\left(-\frac{2\pi}{3}\right)} = \cos\left(-\frac{2\pi}{3}\right) + i\sin\left(-\frac{2\pi}{3}\right) = -\frac{1}{2} - \frac{\sqrt{3}}{2}\,i = z_2$$
$$z_3 = e^{i\frac{6\pi}{3}}\cos(2\pi) + i\sin(2\pi) = 1 = z_0$$

We see that as we increase (resp. decrease) k by 1, we rotate counterclockwise (resp. clockwise) by an angle of $2\pi/3$, and thus after doing so three times, we are back where we started. The 3rd roots of 1 are plotted in Figure 7.5.1.

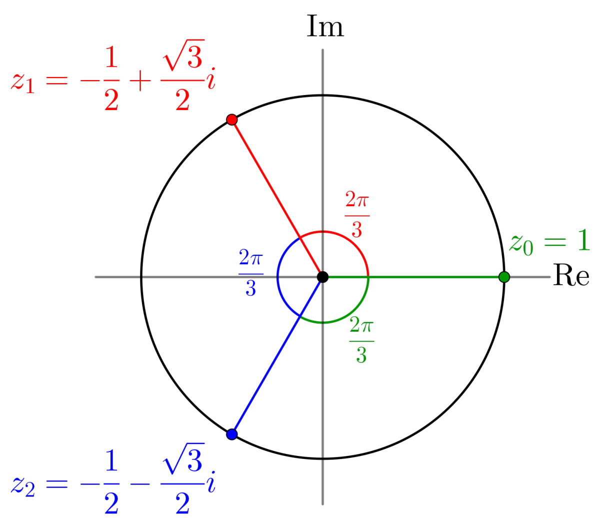


Figure 7.5.1: The 3rd roots of 1.

Example 7.5.4 Find all 4th roots of -256 in standard form and plot them in the complex plane.

Solution: Here, $w = -256$ and $n = 4$. We have that $-256 = 256(\cos\pi + i\sin\pi) = 256e^{i\pi}$ so the 4th roots are given by

$$z_k = (256)^{1/4}e^{i\frac{\pi+2k\pi}{4}} = 4\left(\cos\left(\frac{\pi+2k\pi}{4}\right) + i\sin\left(\frac{\pi+2k\pi}{4}\right)\right),$$

for $k = 0, 1, 2, 3$. Thus

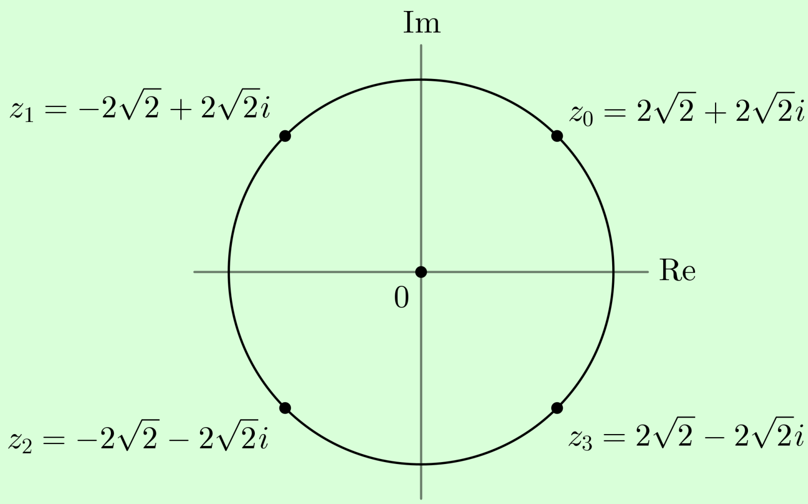
$$z_0 = 4\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right) = 4\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right) = 2\sqrt{2} + 2\sqrt{2}i$$

$$z_1 = 4\left(\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4}\right) = 4\left(-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right) = -2\sqrt{2} + 2\sqrt{2}i$$

$$z_2 = 4\left(\cos\frac{5\pi}{4} + i\sin\frac{5\pi}{4}\right) = 4\left(-\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right) = -2\sqrt{2} - 2\sqrt{2}i$$

$$z_3 = 4\left(\cos\frac{7\pi}{4} + i\sin\frac{7\pi}{4}\right) = 4\left(\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right) = 2\sqrt{2} - 2\sqrt{2}i$$

which we plot in the complex plane. Notice again that the roots are evenly spaced out on a circle of radius 4.



Exercise 122 Find the 3rd roots of $4 - 4\sqrt{3}i$. Express your answers in polar form.

Section 7.5 Problems

7.5.1. Find

(a) all cube roots of $-2+2i$ in complex exponential form.

(b) all sixth roots of $-i$ in polar form.

(c) all fourth roots of -4 in standard form.

7.5.2. Let n be a positive integer. A number $z \in \mathbb{C}$ is called an *nth root of unity* if $z^n = 1$.

(a) Show that ± 1 and $\pm i$ are nth roots of unity by finding a suitable n.

(b) Show that if z is an nth root of unity then $|z| = 1$.

(c) Show that if z is an nth root of unity then so is z^m for any integer m.

(d) Find all 3rd roots of unity in standard form.

Chapter 8

Eigenvalues and Eigenvectors

Our course in linear algebra began with a study of vector geometry and systems of equations, topics we gained a deeper understanding of as we examined matrix algebra. Then we focused on the notions of span, linear independence, subspaces, bases and dimension, which ultimately gave us the framework under which linear algebra operates. By this point, we had stated the Matrix Invertibility Criteria (first as Theorem 3.5.13 and later as Theorem 4.7.1), which tied together the many important concepts we have encountered and served to illustrate how intertwined the topics of linear algebra truly are.

From there, we proceeded to study linear transformations, employing many of the results we had derived in the first part of the course. This was followed by a brief examination of determinants, both as an indicator of invertibility and as a tool to compute areas and volumes. Finally, we focused on complex numbers, a choice that will have likely have felt more like a distraction than an advancement of linear algebra, but we will see soon that it was a necessary diversion.

The topic of this chapter, eigenvalues and eigenvectors, is really the heart of linear algebra. By studying eigenvalues and eigenvectors, we will gain a deeper geometric and algebraic understanding of linear transformations. The concepts discussed in this chapter will draw heavily on *every* topic we have covered thus far: vector geometry, systems of equations, matrix algebra, subspaces, bases, determinants and complex numbers will indeed all make an appearance.

The results of this chapter have applications throughout all of mathematics, science and engineering. Some areas where eigenvalues and eigenvectors will arise include the study of heat transfer, control theory, vibration analysis, the modelling of electric circuits, power system analysis, facial recognition, predator-prey models, quantum mechanics and systems of linear differential equations.

8.1 Introduction

To motivate our work in this chapter, we will consider a couple examples involving reflections and projections.

8.1.1 Example: Reflections Through Lines in $\mathbb{R}^2$

We consider first a reflection of a vector through the x_1-axis given by the transformation $R : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$R(\vec{x}) = 2\,\mathrm{proj}_{\vec{e}_1}\,\vec{x} - \vec{x},$$

which was shown to be a linear transformation in Example 5.2.3. Figure 8.1.1 illustrates this reflection for a vector $\vec{x} \in \mathbb{R}^2$.

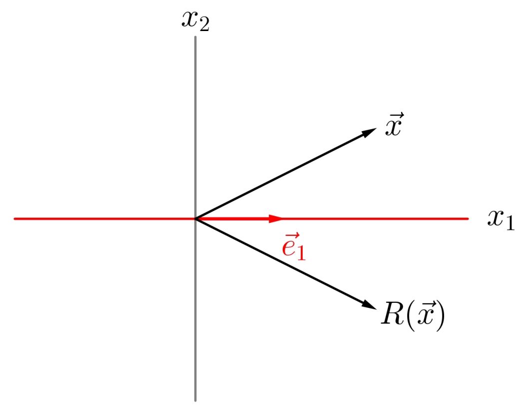


Figure 8.1.1: R is a reflection through the x_1-axis.

Recalling that $\{\vec{e}_1, \vec{e}_2\}$ denotes the standard basis for $\mathbb{R}^2$, we immediately see that

$$\big[R\big] = \big[R(\vec{e}_1) \;\; R(\vec{e}_2)\big] = \big[\,\vec{e}_1 \;\; -\vec{e}_2\big] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

from which it follows that for $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$,

$$R(\vec{x}) = \big[R\big]\,\vec{x} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}.$$

Algebraically, we see that evaluating $R(\vec{x})$ amounts to simply multiplying the second entry of $\vec{x}$ by -1. On the other hand, given only the matrix

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

and defining the matrix transformation $f_A : \mathbb{R}^2 \to \mathbb{R}^2$ by $f_A(\vec{x}) = A\vec{x}$, it should be clear that

$$f_A(\vec{x}) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix} = R(\vec{x}),$$

that is, it should be clear that A is the standard matrix for a reflection in the x_1-axis.

Let us now look at a reflection through a line other than the x_1-axis. Let $\vec{d} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right] \in \mathbb{R}^2$, and consider the line L containing the origin with direction vector $\vec{d}$. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(\vec{x}) = 2\,\mathrm{proj}_{\vec{d}}\,\vec{x} - \vec{x}.$$

Then T is a linear transformation that reflects $\vec{x} \in \mathbb{R}^2$ through the line L as verified in Example 5.2.3 and illustrated in Figure 8.1.2

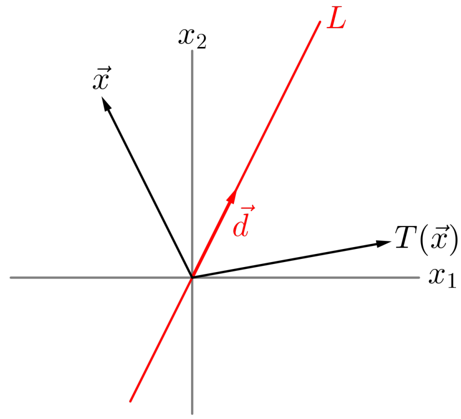


Figure 8.1.2: T is a reflection through the line L.

It is less likely that one can compute $\left[\,T\,\right]$ by inspection, but by using the above definition of T, we arrive at

$$\left[\,T\,\right] = \left[\,T(\vec{e}_1)\;\; T(\vec{e}_2)\,\right] = \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}$$

from which it follows that for $\vec{x} = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right] \in \mathbb{R}^2$,

$$T(\vec{x}) = \left[\,T\,\right]\vec{x} = \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -3x_1 + 4x_2 \\ 4x_1 + 3x_2 \end{bmatrix}.$$

We see that it it is more involved (and thus more error-prone) to determine $\left[\,T\,\right]$ and $T(\vec{x})$ than it is to determine $\left[\,R\,\right]$ and $R(\vec{x})$. Moreover, given just the matrix

$$B = \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix},$$

it is certainly not obvious that the matrix transformation $f_B : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f_B(\vec{x}) = \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -3x_1 + 4x_2 \\ 4x_1 + 3x_2 \end{bmatrix} = T(\vec{x})$$

is a reflection through the line L.

Notice that both R and T are both easily understood geometrically as they are both reflections through lines containing the origin. However, our above work shows that algebraically, it is significantly easier to work with R than T. We now show that this is a result of the standard basis being the "natural" basis for R, but not for T.

Since $\{\vec{e}_1, \vec{e}_2\}$ is the standard basis for $\mathbb{R}^2$, given any $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$, we can write $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2$. Then, since R is linear,

$$R(\vec{x}) = R(x_1\vec{e}_1 + x_2\vec{e}_2) = x_1R(\vec{e}_1) + x_2R(\vec{e}_2) = x_1\vec{e}_1 + x_2(-\vec{e}_2) = x_1\vec{e}_1 - x_2\vec{e}_2 \quad (8.1)$$

Thus, as we have observed before, R takes every linear combination $x_1\vec{e}_1 + x_2\vec{e}_2$ of the vectors $\vec{e}_1$ and $\vec{e}_2$ and simply changes the sign of x_2. This is illustrated in Figure 8.1.3.

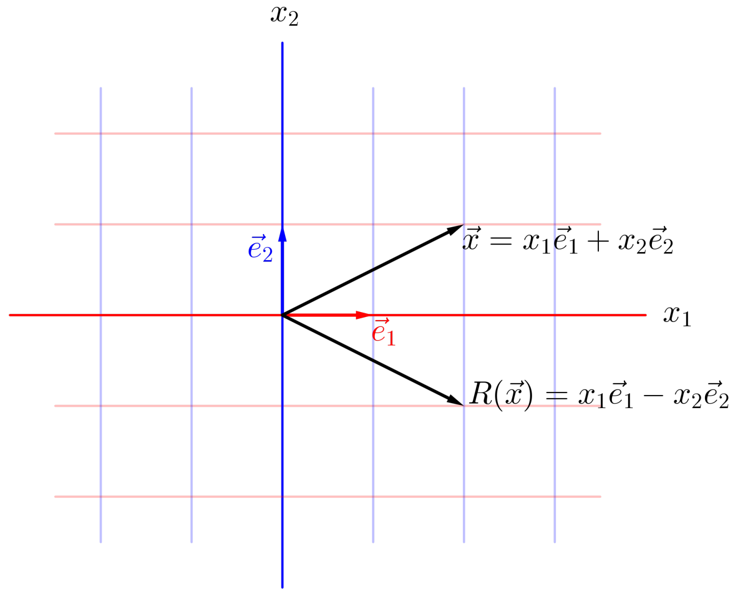


Figure 8.1.3: A vector $\vec{x} \in \mathbb{R}^2$ and its reflection $R(\vec{x})$ in the x_1-axis, both expressed as linear combinations of the standard basis vectors.

However, for T we observe that

$$\begin{aligned} T(\vec{x}) &= T(x_1\vec{e}_1 + x_2\vec{e}_2) \\ &= x_1T(\vec{e}_1) + x_2T(\vec{e}_2) \\ &= x_1\begin{bmatrix} -3/5 \\ 4/5 \end{bmatrix} + x_2\begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} \\ &= \left(-\frac{3}{5}x_1 + \frac{4}{5}x_2\right)\vec{e}_1 + \left(\frac{4}{5}x_1 + \frac{3}{5}x_2\right)\vec{e}_2. \end{aligned} \quad (8.2)$$

We see from (8.1) that $R(\vec{x})$ is naturally expressed in the standard basis whereas (8.2) shows that T is not. A key observation about R from Figure 8.1.1 is that any vector $\vec{x}$ lying on the x_1-axis (the line R is reflecting through) satisfies $R(\vec{x}) = \vec{x} = 1\vec{x}$ and any vector $\vec{x}$ lying on the x_2-axis (the line perpendicular to the x_1-axis) satisfies $R(\vec{x}) = -\vec{x} = (-1)\vec{x}$. That is, for any vector $\vec{x}$ lying on either the x_1-axis or the x_2-axis, $R(\vec{x})$ is simply a scalar multiple of $\vec{x}$. This gives us a starting point to find an appropriate basis for $\mathbb{R}^2$ that will allow us to better understand T.

In Figure 8.1.4, we consider several vectors in $\mathbb{R}^2$ and their images under T, that is, their reflections in the line L (note that the line L' is the line containing the origin that is perpendicular to L). We observe that any vector $\vec{x}$ lying on the line L appears to satisfy $T(\vec{x}) = \vec{x} = 1\vec{x}$, and any vector $\vec{x}$ lying on the line L' appears to satisfy $T(\vec{x}) = -\vec{x} = (-1)\vec{x}$.

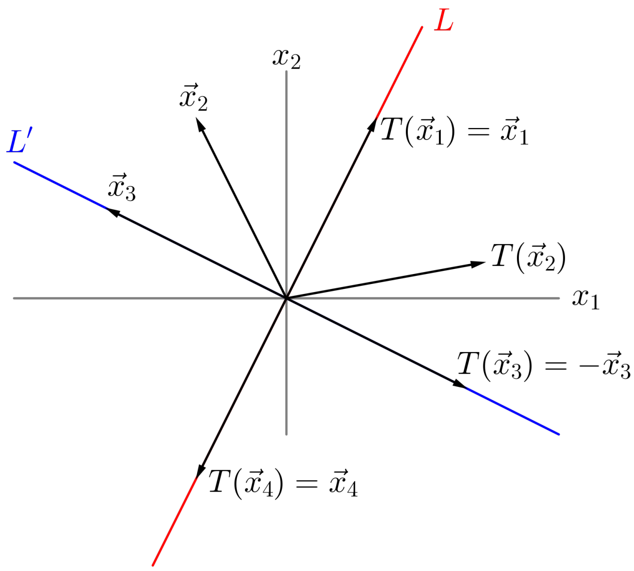


Figure 8.1.4: Plotting $\vec{x}$ and $T(\vec{x})$ for several choices of $\vec{x}$.

Exercise 123 With $\vec{d} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ and $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\vec{x}) = 2\,\text{proj}_{\vec{d}}\,\vec{x} - \vec{x}$, verify that

(a) $T(\vec{x}) = \vec{x}$ for every $\vec{x} \in L$, where L is the line containing the origin with direction vector $\vec{d}$,

(b) $T(\vec{x}) = -\vec{x}$ for every $\vec{x} \in L'$, where L' is the line containing the origin that is perpendicular to L.

In particular, since $\vec{d} = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ is a direction vector for L, we see that $\vec{d} \in L$ and $T(\vec{d}) = \vec{d}$. If we let, say $\vec{n} = \left[\begin{smallmatrix}-2\\1\end{smallmatrix}\right]$, be a direction vector for L', then $\vec{n} \in L'$ and $T(\vec{n}) = -\vec{n}$. Since $\vec{d}$ and $\vec{n}$ are nonzero and not parallel, the set $B = \{\vec{d}, \vec{n}\}$ is linearly independent set of two vectors in $\mathbb{R}^2$. Hence, B is a basis for $\mathbb{R}^2$. Thus, for any $\vec{x} \in \mathbb{R}^2$ we can find $c_1, c_2 \in \mathbb{R}$ so that

$$\vec{x} = c_1\,\vec{d} + c_2\,\vec{n}$$

and since T is linear,

$$T(\vec{x}) = T(c_1\,\vec{d} + c_2\,\vec{n}) = c_1 T(\vec{d}) + c_2 T(\vec{n}) = c_1\,\vec{d} + c_2(-\vec{n}) = c_1\,\vec{d} - c_2\,\vec{n}. \tag{8.3}$$

We see that T takes every linear combination $c_1\,\vec{d} + c_2\,\vec{n}$ of the vectors $\vec{d}$ and $\vec{n}$ and changes the sign of c_2 - in much the same way R does with when working with the standard basis (see (8.1))! This is shown in Figure 8.1.5, which should be compared to Figure 8.1.3.

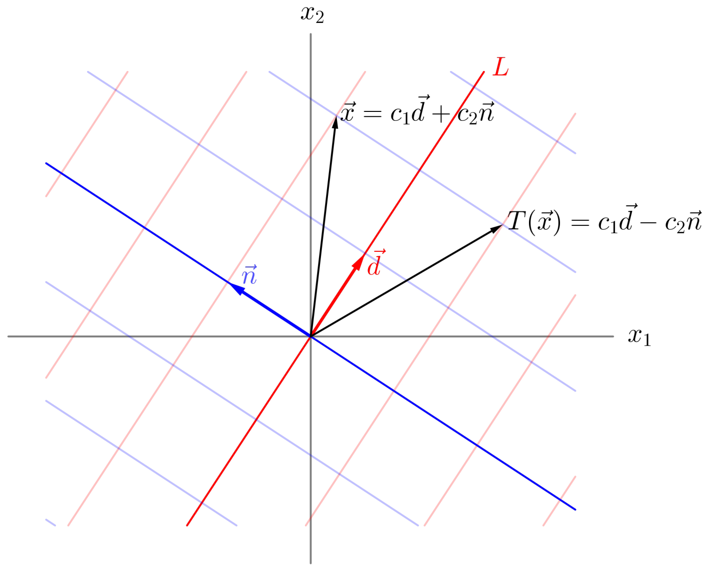


Figure 8.1.5: A vector $\vec{x} \in \mathbb{R}^2$ and its reflection $T(\vec{x})$ in the line L, both expressed as linear combinations of the basis vectors $\vec{d}$ and $\vec{n}$.

Thus, we see that T is more easily understood algebraically when we work in the basis $\{\vec{d}, \vec{n}\}$ rather than the standard basis $\{\vec{e}_1, \vec{e}_2\}$ since the image of every vector $\vec{x} = c_1\vec{d} + c_2\vec{n}$ under T is simply $T(\vec{x}) = c_1\vec{d} - c_2\vec{n}$. In fact, observe that

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ -c_2 \end{bmatrix}.$$

Thus if we consider just the coefficients c_1 and c_2 that express $\vec{x}$ as a linear combination of $\vec{d}$ and $\vec{n}$, then we can use a diagonal matrix to compute the coefficients needed to express $T(\vec{x})$ as a linear combination of $\vec{d}$ and $\vec{n}$.

This understanding of T began by simply trying to find those vectors $\vec{x} \in \mathbb{R}^2$ such that $T(\vec{x})$ is a scalar multiple of $\vec{x}$. This leads to the following important definition.

Definition 8.1.1

Eigenvalue, Eigenvector

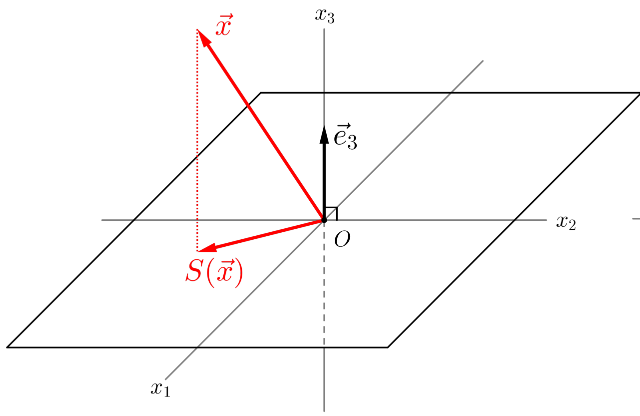
For a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, a scalar λ is an **eigenvalue** of T if $T(\vec{x}) = \lambda\vec{x}$ for some *nonzero* vector $\vec{x}$. The vector $\vec{x}$ is then called an **eigenvector** of T corresponding to λ.

We make a couple of remarks here. First, since we are requiring $T(\vec{x})$ to be a scalar multiple of $\vec{x}$, we must have that T is a linear transformation with $\mathbb{R}^n$ being both the domain and the codomain. Secondly, note that we do not allow for the zero vector to be an eigenvector. This is simply because $T(\vec{0}) = \vec{0}$ for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, meaning that $T(\vec{0}) = \lambda\vec{0}$ is trivially true for any scalar λ, which does not give us any meaningful information.

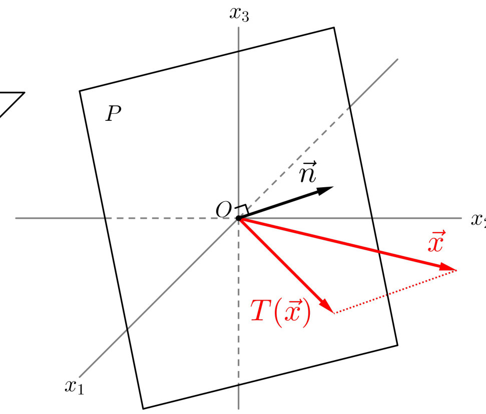
Our work above is easily generalized to any reflection in $\mathbb{R}^2$ through a line L containing the origin with direction vector $\vec{d}$. Simply pick any nonzero vector $\vec{n} \in \mathbb{R}^2$ that is orthogonal to $\vec{d}$ so that the set $\{\vec{d}, \vec{n}\}$ will be a basis for $\mathbb{R}^2$ for which $T(\vec{d}) = \vec{d}$ and $T(\vec{n}) = -\vec{n}$, and (8.3) will then hold.

8.1.2 Example: Projections onto Planes in $\mathbb{R}^3$

Consider the transformations $S, T : \mathbb{R}^3 \to \mathbb{R}^3$ where S is a projection onto the x_1x_2-plane given by the scalar equation $x_3 = 0$, and T is a projection onto the plane P given by the scalar equation $x_1 + 2x_2 + x_3 = 0$. It follows from Exercise 80 that both S and T are linear transformations, and both are illustrated in Figure 8.1.6.



(a) S is a projection onto the x_1x_2-plane.



(b) T is a projection onto the plane P.

Figure 8.1.6: The projections S and T.

Considering S first, we see that $\vec{e}_3$ is a normal vector for the x_1x_2-plane, so we have that $S(\vec{x}) = \vec{x} - \operatorname{proj}_{\vec{e}_3} \vec{x}$. It follows that

$$\big[\,S\,\big] = \big[\,S(\vec{e}_1)\ \ S(\vec{e}_2)\ \ S(\vec{e}_3)\,\big] = \big[\,\vec{e}_1\ \ \vec{e}_2\ \ \vec{0}\,\big] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so for $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right] \in \mathbb{R}^3$, we have

$$S(\vec{x}) = \big[\,S\,\big]\,\vec{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}.$$

Simply put, projecting a vector in $\mathbb{R}^3$ onto the x_1x_2-plane simply changes the x_3-coordinate of that vector to 0. Given the matrix

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

and defining a matrix transformation $f_A : \mathbb{R}^3 \to \mathbb{R}^3$ by $f_A(\vec{x}) = A\vec{x}$, it should also be clear that

$$f_A(\vec{x}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} = S(\vec{x}),$$

that is, it should be clear that A is the standard matrix for a projection onto the x_1x_2-plane.

Turning our attention to T (which we recall is a projection onto the plane P with scalar equation $x_1 + 2x_2 + x_3 = 0$), we see that $\vec{n} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ is a normal vector for P so that $T(\vec{x}) = \vec{x} - \operatorname{proj}_{\vec{n}} \vec{x}$. With a bit of work, we can show that

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) & T(\vec{e}_3) \end{bmatrix} = \begin{bmatrix} 5/6 & -1/3 & -1/6 \\ -1/3 & 1/3 & -1/3 \\ -1/6 & -1/3 & 5/6 \end{bmatrix},$$

and it follows that for $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$,

$$T(\vec{x}) = \begin{bmatrix} T \end{bmatrix} \vec{x} = \begin{bmatrix} 5/6 & -1/3 & -1/6 \\ -1/3 & 1/3 & -1/3 \\ -1/6 & -1/3 & 5/6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{5}{6}x_1 - \frac{1}{3}x_2 - \frac{1}{6}x_3 \\ -\frac{1}{3}x_1 + \frac{1}{3}x_2 - \frac{1}{3}x_3 \\ -\frac{1}{6}x_1 - \frac{1}{3}x_2 + \frac{5}{6}x_3 \end{bmatrix}.$$

Note that it is more difficult to compute $\begin{bmatrix} T \end{bmatrix}$ and $T(\vec{x})$ than it is to compute $\begin{bmatrix} S \end{bmatrix}$ and $S(\vec{x})$. Also note that given the matrix

$$B = \begin{bmatrix} 5/6 & -1/3 & -1/6 \\ -1/3 & 1/3 & -1/3 \\ -1/6 & -1/3 & 5/6 \end{bmatrix},$$

it is not clear that the matrix transformation $f_B(\vec{x}) : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $f_B(\vec{x}) = B\vec{x}$ is a projection onto the plane P.

For any $\vec{x} \in \mathbb{R}^3$, we have that

$$S(\vec{x}) = S(x_1\vec{e}_2 + x_2\vec{e}_2 + x_3\vec{e}_3) = x_1 S(\vec{e}_1) + x_2 S(\vec{e}_2) + x_2 S(\vec{e}_3) = x_1\vec{e}_1 + x_2\vec{e}_2 \quad (8.4)$$

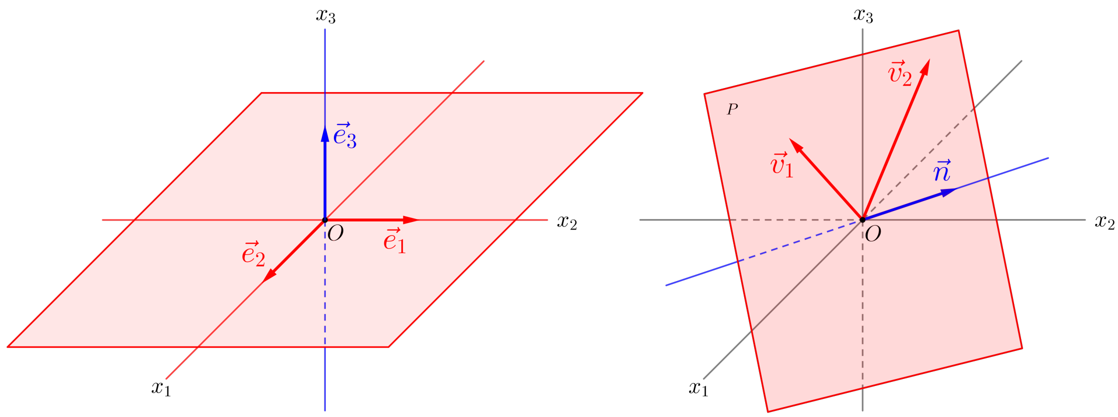
and that

$$\begin{aligned} T(\vec{x}) &= T(x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3) \\ &= x_1 T(\vec{e}_1) + x_2 T(\vec{e}_2) + x_2 T(\vec{e}_3) \\ &= x_1 \begin{bmatrix} 5/6 \\ -1/3 \\ -1/6 \end{bmatrix} + x_2 \begin{bmatrix} -1/3 \\ 1/3 \\ -1/3 \end{bmatrix} + x_3 \begin{bmatrix} -1/6 \\ -1/3 \\ 5/6 \end{bmatrix} \\ &= \left(\frac{5}{6}x_1 - \frac{1}{3}x_2 - \frac{1}{6}x_3\right)\vec{e}_1 + \left(-\frac{1}{3}x_1 + \frac{1}{3}x_2 - \frac{1}{3}x_3\right)\vec{e}_2 + \left(-\frac{1}{6}x_1 - \frac{1}{3}x_2 + \frac{5}{6}x_3\right)\vec{e}_3. \end{aligned} \quad (8.5)$$

It is clear from (8.4) and (8.5) that S is expressed naturally in the standard basis but that T is not. Indeed, from above, we see that

$$S(\vec{e}_1) = 1\vec{e}_1, \quad S(\vec{e}_2) = 1\vec{e}_2 \quad \text{and} \quad S(\vec{e}_3) = 0\vec{e}_3.$$

It follows that $\lambda_1 = 1$ is an eigenvalue of S with corresponding eigenvectors $\vec{e}_1$ and $\vec{e}_2$, and that $\lambda_2 = 0$ is an eigenvalue of S with corresponding eigenvector $\vec{e}_3$. We notice that $\{\vec{e}_1, \vec{e}_2\}$ forms a basis for the x_1x_2-plane and $\{\vec{e}_3\}$ is a basis for x_3-axis (the line through the origin that is perpendicular to the x_1x_2-plane), and that together, $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ forms



(a) Every nonzero vector in the x_1x_2-plane is an eigenvector of S corresponding to the eigenvalue $\lambda_1 = 1$. Every nonzero vector on the x_3-axis is an eigenvector of S corresponding to the eigenvalue $\lambda_2 = 0$. The set $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ is a basis for $\mathbb{R}^3$ consisting of eigenvectors of S.

(b) Every nonzero vector in the plane P is an eigenvector of T corresponding to the eigenvalue $\mu_1 = 1$. Every nonzero vector on the line containing the origin with direction vector $\vec{n}$ is an eigenvector of T corresponding to the eigenvalue $\mu_2 = 0$. The set $\{\vec{v}_1, \vec{v}_2, \vec{n}\}$ is a basis for $\mathbb{R}^3$ consisting of eigenvectors of T.

Figure 8.1.7: A projection onto any plane $P \subseteq \mathbb{R}^3$ will not "move" any vector in P, and will "send" any scalar multiple of a normal vector of P to the zero vector.

a basis for $\mathbb{R}^3$ consisting of eigenvectors of S (in this case, the standard basis). See Figure 8.1.7a.

We now construct a basis for $\mathbb{R}^3$ consisting of eigenvectors of T in a similar way. Let $\{\vec{v}_1, \vec{v}_2\}$ be any basis for P. Since $\vec{v}_1, \vec{v}_2 \in P$, we have that

$$T(\vec{v}_1) = 1\vec{v}_1 \quad \text{and} \quad T(\vec{v}_2) = 1\vec{v}_2$$

so $\mu_1 = 1$ is an eigenvalue of T with corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$. Let $\{\vec{n}\}$ be a basis for the line through the origin that is perpendicular to P. Then $T(\vec{n}) = \vec{0}$, that is,

$$T(\vec{n}) = 0\vec{n}$$

so $\mu_2 = 0$ is an eigenvalue[1] of T with a corresponding eigenvector $\vec{n}$. Taken together, it is not difficult to see that the set $\{\vec{v}_1, \vec{v}_2, \vec{n}\}$ forms a basis for $\mathbb{R}^3$. This is illustrated in Figure 8.1.7b

Exercise 124 Let $P \subseteq \mathbb{R}^3$ be a plane with scalar equation $x_1 + 2x_2 + x_3 = 0$ and let $B = \{\vec{v}_1, \vec{v}_2\}$ be a basis for P and let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a projection onto P.

(a) Verify algebraically that $\vec{v}_1$ and $\vec{v}_2$ are eigenvectors of T corresponding to the eigenvalue $\mu_1 = 1$.

(b) Show that every nonzero linear combination of $\vec{v}_1$ and $\vec{v}_2$ is also an eigenvector of T corresponding to $\mu_1 = 1$.

[1] Note that Definition 8.1.1 excludes the vector $\vec{0}$ from being an *eigenvector* of T, but it does not exclude the number 0 from being an *eigenvalue* of T.

(c) Verify algebraically that $\vec{n} = \left[\begin{smallmatrix}1\\2\\1\end{smallmatrix}\right]$ is an eigenvector of T corresponding to the eigenvalue $\mu_2 = 0$.

(d) Show that every nonzero linear combination of $\vec{n}$ is also a eigenvector of T corresponding to $\mu_2 = 0$.

Exercise 125 Let $P \subseteq \mathbb{R}^3$ be the plane with scalar equation $x_1 + 2x_2 + x_3 = 0$. Find a basis for P.

Finally, for any $\vec{x} \in \mathbb{R}^3$, there are $c_1, c_2, c_3 \in \mathbb{R}$ so that $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{n}$. We thus have

$$T(\vec{x}) = T(c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{n}) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + c_3T(\vec{n}) = c_1\vec{v}_1 + c_2\vec{v}_2. \tag{8.6}$$

We see that when working with the basis $\{\vec{v}_1, \vec{v}_2, \vec{n}\}$, T behaves exactly like S does when working with the standard basis (see (8.5)). Note that

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}c_1\\c_2\\0\end{bmatrix},$$

so if we consider just the coefficients c_1, c_2 and c_3 that express $\vec{x}$ as a linear combination of $\vec{v}_1$, $\vec{v}_2$ and $\vec{n}$, then we can use a diagonal matrix to compute the coefficients needed to express $T(\vec{x})$ as a linear combination of $\vec{v}_1$, $\vec{v}_2$ and $\vec{n}$.

Again, the work we have done above generalizes to a projection onto any plane P in $\mathbb{R}^3$ containing the origin. If we take $\{\vec{v}_1, \vec{v}_2\}$ to be a basis for P and $\vec{n}$ to be a normal vector for P, then $\{\vec{v}_1\vec{v}_2, \vec{n}\}$ will be a basis for $\mathbb{R}^3$ consisting of eigenvectors of T, and (8.6) will hold.

Section 8.1 Problems

8.1.1. Consider two linear transformations $S, T : \mathbb{R}^2 \to \mathbb{R}^2$, both of which are projections onto lines in $\mathbb{R}^2$ (see Example 5.2.2): S is the projection onto the x_1-axis and T is the projection onto the line L containing the origin with direction vector $\vec{d} = \left[\begin{smallmatrix}2\\3\end{smallmatrix}\right]$.

(a) Determine $\left[S\right]$.

(b) For $\vec{x} = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]$, express $S(\vec{x})$ as a linear combination of $\vec{e}_1$ and $\vec{e}_2$.

(c) Determine $\left[T\right]$.

(d) For $\vec{x} = \left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]$, express $T(\vec{x})$ as a linear combination of $\vec{e}_1$ and $\vec{e}_2$.

(e) Geometrically determine the eigenvalues and corresponding eigenvectors of T.

(f) Determine a basis B for $\mathbb{R}^2$ that consists of eigenvectors of T.

(g) For $\vec{x} \in \mathbb{R}^2$, express $T(\vec{x})$ as a linear combination of the basis vectors in B. Compare your result to part (b).

8.1.2. Consider two linear transformations $R, T : \mathbb{R}^3 \to \mathbb{R}^3$, both of which are reflections through planes in $\mathbb{R}^3$ (see Example 5.2.4): R is the reflection through the plane $x_3 = 0$ and T is the reflection through the plane $x_1 + x_2 + x_3 = 0$.

(a) Determine $\left[R\right]$.

(b) For $\vec{x} = \left[\begin{smallmatrix}x_1\\x_2\\x_3\end{smallmatrix}\right]$, express $R(\vec{x})$ as a linear combination of $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$.

(c) Determine $\left[T\right]$.

(d) For $\vec{x} = \left[\begin{smallmatrix}x_1\\x_2\\x_3\end{smallmatrix}\right]$, express $T(\vec{x})$ as a linear combination of $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$.

(e) Geometrically determine the eigenvalues and corresponding eigenvectors of T.

(f) Determine a basis B for $\mathbb{R}^3$ that consists of eigenvectors of T.

(g) For $\vec{x} \in \mathbb{R}^3$, express $T(\vec{x})$ as a linear combination of the vectors in B. Compare your result to part (b).

8.2 Computing Eigenvalues and Eigenvectors

The examples presented in Section 8.1 relied on our having some geometric intuition about a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ so that we could generate eigenvalues λ and corresponding eigenvectors $\vec{x} \in \mathbb{R}^n$ so that $T(\vec{x}) = \lambda\vec{x}$. However, we won't always be able to find the eigenvalues and corresponding eigenvectors of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ in this way. For example, the linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^4$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} 5x_1 - 23x_2 + 18x_3 - 102x_4 \\ -23x_1 + 14x_2 - 6x_3 + 73x_4 \\ 123x_1 + 34x_4 \\ x_1 + x_2 - x_3 + 56x_4 \end{bmatrix}$$

is not easily understood geometrically, so it becomes difficult to find eigenvalues and the corresponding eigenvectors of T using the methods of Section 8.1. In this section, we will derive an algebraic technique that does not rely on the geometry of a linear transformation to determine eigenvalues and eigenvectors. This method will focus on the standard matrix of a linear transformation, so we make the following definition which is the "matrix analogue" of Definition 8.1.1.

Definition 8.2.1 **Eigenvalue, Eigenvector**

For $A \in M_{n\times n}(\mathbb{R})$, a scalar λ is an **eigenvalue** of A if $A\vec{x} = \lambda\vec{x}$ for some *nonzero* vector $\vec{x}$. The vector $\vec{x}$ is then called an **eigenvector** of A corresponding to λ.

We begin with a couple of straightforward examples to ensure we understand this definition.

Example 8.2.2

If $A = \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then

$$A\vec{x} = \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1\vec{x},$$

and so $\lambda = 1$ is an eigenvalue of A and $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is a corresponding eigenvector.

Note that the matrix A in Example 8.2.2 is the standard matrix of the linear transformation T from Subsection 8.1.1.

Example 8.2.3

If $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ then

$$A\vec{x} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0\vec{x}$$

and so $\lambda = 0$ is an eigenvalue of A and $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a corresponding eigenvector.

Example 8.2.3 shows us that a matrix *can* have $\lambda = 0$ as an eigenvalue. However, it *can never* have $\vec{x} = \vec{0}$ as an eigenvector because according to Definition 8.2.1, eigenvectors must be nonzero.

Exercise 126 Find an eigenvalue and a corresponding eigenvector for $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

We now look at how to algebraically determine the eigenvalues and corresponding eigenvectors for a matrix $A \in M_{n\times n}(\mathbb{R})$. Definition 8.2.1 states that a scalar λ is an eigenvalue of A with corresponding eigenvector $\vec{x} \neq \vec{0}$ if and only if

$$\begin{aligned} A\vec{x} = \lambda\vec{x} &\iff A\vec{x} - \lambda\vec{x} = \vec{0} \\ &\iff A\vec{x} - \lambda I\vec{x} = \vec{0} \quad \text{since } I\vec{x} = \vec{x} \\ &\iff (A - \lambda I)\vec{x} = \vec{0}. \end{aligned}$$

Thus we will consider the homogeneous system $(A - \lambda I)\vec{x} = \vec{0}$. Since $\vec{x} \neq \vec{0}$, we require nontrivial solutions to this system, and since $A - \lambda I \in M_{n\times n}(\mathbb{R})$, the Matrix Invertibility Criteria Revisited gives that $A - \lambda I$ cannot be invertible. It then follows from Theorem 6.1.11 that $\det(A - \lambda I) = 0$. This verifies the following theorem.

Theorem 8.2.4 Let $A \in M_{n\times n}(\mathbb{R})$. A scalar λ is an eigenvalue of A if and only if λ satisfies the equation

$$\det(A - \lambda I) = 0.$$

If λ is a eigenvalue of A, then all nontrivial solutions of the homogeneous system of equations

$$(A - \lambda I)\vec{x} = \vec{0}$$

are the eigenvectors of A corresponding to λ.

Theorem 8.2.4 indicates that finding the eigenvalues and corresponding eigenvectors of a matrix $A \in M_{n\times n}(\mathbb{R})$ is a two-step process: we first find the eigenvalues λ of A by solving $\det(A - \lambda I) = 0$, and then for each eigenvalue λ of A, we find the corresponding eigenvectors by solving $(A - \lambda I)\vec{x} = \vec{0}$.

We focus first on finding the eigenvalues of a matrix $A \in M_{n\times n}(\mathbb{R})$.

Example 8.2.5 Let $A = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}$. Find the eigenvalues of A.

Solution: We have

$$\begin{aligned} \det(A - \lambda I) = \det\left(\begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) &= \begin{vmatrix} 1-\lambda & 3 \\ 4 & 5-\lambda \end{vmatrix} \\ &= (1-\lambda)(5-\lambda) - 12 \\ &= \lambda^2 - 6\lambda - 7 \\ &= (\lambda + 1)(\lambda - 7). \end{aligned}$$

From this, we see that $\det(A - \lambda I) = 0$ if and only if $\lambda = -1$ or $\lambda = 7$. Thus $\lambda_1 = -1$ and $\lambda_2 = 7$ are the eigenvalues of A.

Note that when a matrix has multiple eigenvalues, we normally list them as $\lambda_1, \lambda_2, \ldots$. It does not matter the order that you do this in - we could have given the solution to Example 8.2.5 as $\lambda_1 = 7$ and $\lambda_2 = -1$.

Exercise 127 Find the eigenvalues of $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

We notice from Example 8.2.5 and Exercise 127 that $\det(A - \lambda I)$ is a polynomial. In fact, for $A \in M_{n\times n}(\mathbb{R})$, $\det(A - \lambda I)$ is a real polynomial of degree n (a fact we will not prove). This leads us to make the following definition.

Definition 8.2.6

Characteristic Polynomial

Let $A \in M_{n\times n}(\mathbb{R})$. The **characteristic polynomial** of A is

$$C_A(\lambda) = \det(A - \lambda I).$$

It is immediately clear that λ is an eigenvalue of A if and only if $C_A(\lambda) = 0$.

We now look at finding the eigenvectors that correspond to the eigenvalues of a matrix $A \in M_{n\times n}(\mathbb{R})$.

Example 8.2.7 Let $A = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}$. For each eigenvalue of A, find the corresponding eigenvectors.

Solution: From Example 8.2.5, we have that $\lambda_1 = -1$ and $\lambda_2 = 7$ are the eigenvalues of A. For $\lambda_1 = -1$, we solve $(A - (-1)I)\vec{x} = \vec{0}$, that is, we solve $(A + I)\vec{x} = \vec{0}$:

$$A + I = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} \underset{R_2 - 2R_1}{\longrightarrow} \begin{bmatrix} 2 & 3 \\ 0 & 0 \end{bmatrix} \overset{\frac{1}{2}R_1}{\longrightarrow} \begin{bmatrix} 1 & 3/2 \\ 0 & 0 \end{bmatrix}.$$

We see that

$$\vec{x} = \begin{bmatrix} -3t/2 \\ t \end{bmatrix} = t\begin{bmatrix} -3/2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R},$$

so eigenvectors of A corresponding to $\lambda_1 = -1$ are

$$\vec{x} = t\begin{bmatrix} -3/2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}, t \neq 0.$$

For $\lambda_2 = 7$, we solve $(A - 7I)\vec{x} = \vec{0}$:

$$A - 7I = \begin{bmatrix} -6 & 3 \\ 4 & -2 \end{bmatrix} \underset{R_2 + \frac{2}{3}R_1}{\longrightarrow} \begin{bmatrix} -6 & 3 \\ 0 & 0 \end{bmatrix} \overset{\frac{1}{6}R_1}{\longrightarrow} \begin{bmatrix} 1 & -1/2 \\ 0 & 0 \end{bmatrix}.$$

We have that

$$\vec{x} = \begin{bmatrix} s/2 \\ s \end{bmatrix} = s \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}, \quad s \in \mathbb{R},$$

so the eigenvalues of A corresponding to $\lambda_2 = 7$ are

$$\vec{x} = s \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}, \quad s \in \mathbb{R}, \, s \neq 0.$$

We make a couple of remarks regarding Example 8.2.7. First we see that the eigenvectors corresponding to an eigenvalue λ of A are simply the nontrivial (nonzero) solutions to the homogeneous system $(A - \lambda I)\vec{x} = \vec{0}$.

Secondly, we note that we can scale the vectors $\begin{bmatrix} -3/2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$ by a factor of 2 when finding the eigenvectors corresponding to the eigenvalues of A (see the discussion following Example 2.4.2). This is often done to eliminate fractions in our final answers and can be helpful in Section 8.4. Thus, it is also correct to conclude that

$$\vec{x} = t \begin{bmatrix} -3 \\ 2 \end{bmatrix}, \quad t \in \mathbb{R}, \, t \neq 0 \qquad \text{and} \qquad \vec{x} = s \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad s \in \mathbb{R}, \, s \neq 0$$

are the eigenvectors of A corresponding to $\lambda_1 = -1$ and $\lambda_2 = 7$, respectively.

Exercise 128

For each eigenvalue of $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, find the corresponding eigenvectors.

[**Hint**: Remember that you computed the eigenvalues of A in Exercise 127]

Example 8.2.8

Let $A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Find the eigenvalues of A, and for each eigenvalue find the corresponding eigenvectors.

Solution: We have

$$0 = C_A(\lambda) = \begin{vmatrix} 2-\lambda & 0 & -1 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (2-\lambda)^2(1-\lambda) = -(\lambda-2)^2(\lambda-1)$$

from which we immediately see that $\lambda_1 = 1$ and $\lambda_2 = 2$ are the eigenvalues of A. For $\lambda_1 = 1$, we solve $(A - I)\vec{x} = \vec{0}$:

$$A - I = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus the eigenvectors of A corresponding to $\lambda_1 = 1$ are

$$\vec{x} = \begin{bmatrix} t \\ 0 \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}, \, t \neq 0.$$

For $\lambda_2 = 2$, we solve $(A - 2I)\vec{x} = \vec{0}$:

$$A - 2I = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \underset{R_3 - R_1}{\longrightarrow} \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \overset{-R_1}{\longrightarrow} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus the eigenvectors of A corresponding to $\lambda_2 = 2$ are

$$\vec{x} = \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad s, t \in \mathbb{R},\ s, t \text{ not both zero.}$$

We again make a couple of remarks regarding Example 8.2.8. First, notice that we obtained only two eigenvalues despite A being a 3×3 matrix. Also notice that when solving for the eigenvectors corresponding to $\lambda_2 = 2$, the solution to homogeneous system $(A - 2I)\vec{x} = \vec{0}$ contained two parameters. We will say more about this in Section 8.3.

Secondly, the matrix A is upper triangular from which it follows that the matrix $A - \lambda I$ is also upper triangular. Thus given an upper (or lower) triangular matrix $A \in M_{n\times n}(\mathbb{R})$, the characteristic polynomial $C_A(\lambda)$ of A will be the product of the terms on the main diagonal of $A - \lambda I$ from which it follows that the eigenvalues of A will be the entries lying on the main diagonal of A. This is stated in the following theorem.

Theorem 8.2.9 Let $A \in M_{n\times n}(\mathbb{R})$ be an upper or lower triangular matrix. Then the eigenvalues of A are the entries lying on the main diagonal of A.

Exercise 129 Verify your results from Exercise 126 by finding the eigenvalues and corresponding eigenvectors for $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ using Theorem 8.2.4.

In this section, we have considered matrices $A \in M_{n\times n}(\mathbb{R})$ rather than linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$. The next theorem shows that we have ultimately been computing eigenvalues and eigenvectors of linear transformations.

Theorem 8.2.10 Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, and let $A = \begin{bmatrix} T \end{bmatrix} \in M_{n\times n}(\mathbb{R})$ be the standard matrix of T. Then a scalar λ is an eigenvalue of T with corresponding eigenvector $\vec{x}$ if and only if λ is an eigenvalue of A with corresponding eigenvector $\vec{x}$

Proof: Our result follows immediately from the fact that

$$T(\vec{x}) = \lambda\vec{x} \iff \begin{bmatrix} T \end{bmatrix} \vec{x} = \lambda\vec{x} \iff A\vec{x} = \lambda\vec{x}. \qquad \square$$

Thus, if we are unable to geometrically determine the eigenvalues and corresponding eigenvectors of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, then we can instead algebraically find the eigenvalues and corresponding eigenvectors of $A = \begin{bmatrix} T \end{bmatrix} \in M_{n\times n}(\mathbb{R})$. Additionally, if we have determined the eigenvalues and corresponding eigenvectors of a matrix $A \in M_{n\times n}(\mathbb{R})$, then we have determined the eigenvalues and corresponding eigenvectors of the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(\vec{x}) = f_A(\vec{x}) = A\vec{x}$.

The next example shows that the eigenvalues of $A \in M_{n\times n}(\mathbb{R})$ need not be real.

Example 8.2.11 Find the eigenvalues for $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, and for each eigenvalue, find the corresponding eigenvectors.

Solution: Since

$$0 = C_A(\lambda) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1,$$

we see that $\lambda_1 = i$ and $\lambda_2 = -i$ are the (complex) eigenvalues of A. For $\lambda_1 = i$, we have

$$A - iI = \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \xrightarrow{iR_1} \begin{bmatrix} 1 & -i \\ 1 & -i \end{bmatrix} \underset{R_2-R_1}{\longrightarrow} \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}$$

and we thus see that the (complex) eigenvectors of A corresponding to $\lambda_1 = i$ are

$$\vec{x} = \begin{bmatrix} it \\ t \end{bmatrix} = t\begin{bmatrix} i \\ 1 \end{bmatrix}, \quad t \in \mathbb{C},\ t \neq 0.$$

For $\lambda_2 = -i$, we have

$$A + iI = \begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \xrightarrow{-iR_1} \begin{bmatrix} 1 & i \\ 1 & i \end{bmatrix} \underset{R_2-R_1}{\longrightarrow} \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}$$

which gives that the (complex) eigenvalues of A corresponding to $\lambda_2 = -i$ are

$$\vec{x} = \begin{bmatrix} -it \\ t \end{bmatrix} = t\begin{bmatrix} -i \\ 1 \end{bmatrix}, \quad t \in \mathbb{C},\ t \neq 0.$$

Since the matrix A in Example 8.2.11 has real entries, the characteristic polynomial $C_A(\lambda)$ is a real polynomial as discussed just before Definition 8.2.6. We saw in Section 7.4 that real polynomials may have non-real (complex) roots, but the Conjugate Root Theorem guarantees that that these non-real roots come in "conjugate pairs". Indeed, the roots of $C_A(\lambda)$ were found to be $\lambda_1 = i$ and $\lambda_2 = -i$ which are complex conjugates of one another. Notice that when stating the corresponding eigenvectors for each complex eigenvalue of A, we used complex parameters rather than real parameters.

We also performed elementary row operations on a complex matrix in Example 8.2.11. Many of the results we have derived in this course for matrices $A \in M_{m\times n}(\mathbb{R})$ also hold for matrices $A \in M_{m\times n}(\mathbb{C})$, that is, for $m \times n$ matrices with complex entries. For example, the notions of row echelon and reduced row echelon form generalize naturally to complex matrices and we may use elementary row operations to carry complex matrices to these forms. A second course in linear algebra will explore many of the concepts covered in this course using complex numbers.

We finally note that the matrix A in Example 8.2.11 has a familiar geometric interpretation. Let $R_{\frac{\pi}{2}} : \mathbb{R}^2 \to \mathbb{R}^2$ be a counterclockwise rotation about the origin by an angle of $\pi/2$, which was shown to be a linear transformation in Example 5.2.5. The standard matrix of $R_{\frac{\pi}{2}}$ is

$$\left[R_{\frac{\pi}{2}}\right] = \begin{bmatrix} \cos(\pi/2) & -\sin(\pi/2) \\ \sin(\pi/2) & \cos(\pi/2) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = A.$$

In light of this, it is reasonable that there are no real eigenvalues for A since for any nonzero vector $\vec{x} \in \mathbb{R}^2$, we have that $\vec{x}$ and $A\vec{x}$ are orthogonal and thus $A\vec{x}$ cannot be a scalar multiple of $\vec{x}$.

Exercise 130 Let

$$A = \left[R_\theta\right] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Find all $\theta \in [0, 2\pi)$ such that the eigenvalues of A are real.

Section 8.2 Problems

8.2.1. Find the eigenvalues of A, and for each eigenvalue, find the corresponding eigenvectors.

(a) $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix}$.

(b) $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

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

(d) $A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

8.2.2. Let $A \in M_{n\times n}(\mathbb{R})$ with characteristic polynomial $C_A(\lambda)$.

(a) Prove that $C_{A^T}(\lambda) = C_A(\lambda)$.

(b) Prove that if $r \in \mathbb{R}$ is nonzero, then $C_{rA}(\lambda) = r^n\, C_A\left(\frac{\lambda}{r}\right)$.

8.2.3. Let $A \in M_{n\times n}(\mathbb{R})$.

(a) Prove that if λ is an eigenvalue of A, then $r\lambda$ is an eigenvalue of rA

(b) Assume $A \in M_{3\times 3}(\mathbb{R})$ has eigenvalues $\lambda_1 = 1$, $\lambda_2 = -3$ and $\lambda_3 = 4$. What are the eigenvalues of $8A$?

8.2.4. Let $A \in M_{n\times n}(\mathbb{R})$.

(a) Prove that A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A.

(b) Prove that if A is invertible and λ is an eigenvalue of A, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}.

(c) Let $A \in M_{3\times 3}(\mathbb{R})$ have eigenvalues $\lambda_1 = 2$, $\lambda_2 = -3$ and $\lambda_3 = \frac{2}{3}$. What are the eigenvalues of A^{-1}?

8.2.5. Let $A \in M_{n\times n}(\mathbb{R})$ and let λ be an eigenvalue of A.

(a) Prove that λ^2 is an eigenvalue of A^2.

(b) Prove that λ^3 is an eigenvalue of A^3.

(c) Prove that if λ^m is an eigenvalue of A^m for *some* integer $m \geq 1$, then λ^{m+1} is an eigenvalue of A^{m+1}.

(d) Briefly explain why this shows that λ^k is an eigenvalue of A^k for *every* integer $k \geq 1$. [**Hint**: Use the fact that λ is an eigenvalue of A and think about part (c) when $m = 1$, then when $m = 2$, and so on.]

8.3 Eigenspaces

Given a matrix $A \in M_{n\times n}(\mathbb{R})$, Theorem 8.2.4 tells us that the eigenvalues of A are the roots of the characteristic polynomial $C_A(\lambda) = \det(A - \lambda I)$, and that for each eigenvalue λ of A, we determine the corresponding eigenvalues of A by finding the nontrivial solutions to the homogeneous linear system of equations $(A - \lambda I)\vec{x} = \vec{0}$.

We have seen in Example 8.2.11 that even though $A \in M_{n\times n}(\mathbb{R})$, an eigenvalue λ of A can be non-real and consequently, the eigenvectors of A corresponding to λ can contain non-real entries. In this section, we will study those matrices $A \in M_{n\times n}(\mathbb{R})$ that have only real eigenvalues. Most of the results derived here can be extended to the case when A has non-real eigenvalues in a natural way.

Recall from Definition 4.6.1 that the set of solutions to $(A-\lambda I)\vec{x} = \vec{0}$ is $\mathrm{Null}(A-\lambda I)$, the nullspace of $A - \lambda I$. We make the following definition.

Definition 8.3.1
Eigenspace

Let $\lambda \in \mathbb{R}$ be an eigenvalue of $A \in M_{n\times n}(\mathbb{R})$. The **eigenspace of A corresponding to λ** is the set

$$E_\lambda(A) = \mathrm{Null}(A - \lambda I).$$

Thus, the eigenspace $E_\lambda(A)$ of $A \in M_{n\times n}(\mathbb{R})$ is the set of all eigenvectors of A corresponding to the eigenvalue $\lambda \in \mathbb{R}$ together with the zero vector. Note that if A is the standard matrix of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, that is, if $A = \big[\,T\,\big]$, then we may write $E_\lambda\left(\big[\,T\,\big]\right)$ or $E_\lambda(T)$ instead of $E_\lambda(A)$.

Since $A - \lambda I \in M_{n\times n}(\mathbb{R})$, Theorem 4.6.3 guarantees that $\mathrm{Null}(A - \lambda I)$ is a subspace of $\mathbb{R}^n$. This proves the following result.

Theorem 8.3.2

Let $\lambda \in \mathbb{R}$ be an eigenvalue of $A \in M_{n\times n}(\mathbb{R})$. The eigenspace $E_\lambda(A)$ a subspace of $\mathbb{R}^n$.

Since the eigenspace of a matrix $A \in M_{n\times n}(\mathbb{R})$ corresponding to an eigenvalue $\lambda \in \mathbb{R}$ is a subspace of $\mathbb{R}^n$, we may derive a basis for $E_\lambda(A)$ rather than writing out the vector equation for the solution of $(A - \lambda I)\vec{x} = \vec{0}$ as we did in Section 8.2.

Example 8.3.3

For each eigenvalue λ of $A = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}$, find a basis for $E_\lambda(A)$ and state the dimension of $E_\lambda(A)$.

Solution: From Example 8.2.5, we computed $C_A(\lambda) = (\lambda+1)(\lambda-7)$ from which we deduced that $\lambda_1 = -1$ and $\lambda_2 = 7$ are the eigenvalues of A. From Example 8.2.7, we found that the solution to $(A + I)\vec{x} = \vec{0}$ is

$$\vec{x} = t\begin{bmatrix} -3/2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

so

$$B_1 = \left\{ \begin{bmatrix} -3/2 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_1}(A) = \mathrm{Null}(A+I)$ and $\dim(E_{\lambda_1}(A)) = 1$. Also from Example 8.2.7, the solution to $(A-7I)\vec{x} = \vec{0}$ is

$$\vec{x} = t\begin{bmatrix} 1/2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

so

$$B_2 = \left\{ \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_2}(A) = \mathrm{Null}(A-7I)$ and $\dim(E_{\lambda_2}(A)) = 1$.

Example 8.3.4

For each eigenvalue λ of $A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, find a basis for $E_\lambda(A)$ and state the dimension of $E_\lambda(A)$.

Solution: In Example 8.2.8, we computed $C_A(\lambda) = -(\lambda-2)^2(\lambda-1)$ from which we deduced that the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 2$. We also saw that the solution to $(A-I)\vec{x} = \vec{0}$ is

$$\vec{x} = t\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

so

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_1}(A)$ and $\dim(E_{\lambda_1}(A)) = 1$. We also computed the solution to $(A-2I)\vec{x} = \vec{0}$ as

$$\vec{x} = s\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad s,t \in \mathbb{R},$$

which gives us that

$$B_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_2}(A)$ and $\dim(E_{\lambda_2}(A)) = 2$.

Exercise 131

For each eigenvalue λ of $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, find a basis for $E_\lambda(A)$ and state the dimension of $E_\lambda(A)$. [**Hint**: See Exercise 128.]

In Example 8.3.4 (and thus Example 8.2.8), we see that the eigenvalue $\lambda_2 = 2$ is a repeated root of $C_A(\lambda)$. This motivates the following definition.

Definition 8.3.5 **Algebraic Multiplicity**

Let $A \in M_{n\times n}(\mathbb{R})$ with eigenvalue $\lambda \in \mathbb{R}$. The **algebraic multiplicity** of λ, denoted by a_λ, is the number of times λ appears as a root of $C_A(\lambda)$.

We can determine the algebraic multiplicities of the eigenvalues of a matrix $A \in M_{n\times n}(\mathbb{R})$ by looking at the factorization of the $C_A(\lambda)$.

Example 8.3.6

From Example 8.3.4, the matrix

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

has characteristic polynomial $C_A(\lambda) = -(\lambda-2)^2(\lambda-1)$ and the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 2$. The exponent "1" on the $\lambda-1$ term means that $\lambda_1 = 1$ has algebraic multiplicity 1 and the exponent "2" on the $\lambda - 2$ term means that $\lambda_2 = 2$ has algebraic multiplicity 2. Thus

$$a_{\lambda_1} = 1 \quad \text{and} \quad a_{\lambda_2} = 2.$$

Exercise 132

Find the eigenvalues of

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

For each eigenvalue λ, determine the algebraic multiplicity a_λ.

Given an eigenvalue $\lambda \in \mathbb{R}$ of a matrix $A \in M_{n\times n}(\mathbb{R})$, we will also be concerned with the dimension of the resulting eigenspace, $E_\lambda(A)$. This leads to another definition.

Definition 8.3.7 **Geometric Multiplicity**

Let $A \in M_{n\times n}(\mathbb{R})$ with eigenvalue $\lambda \in \mathbb{R}$. The **geometric multiplicity** of λ, denoted by g_λ, is the dimension of the corresponding eigenspace $E_\lambda(A)$.

Example 8.3.8

Returning to Example 8.3.4, the matrix

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

has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$. We saw that $\dim(E_{\lambda_1}(A)) = 1$ and $\dim(E_{\lambda_2}(A)) = 2$. Thus

$$g_{\lambda_1} = 1 \quad \text{and} \quad g_{\lambda_2} = 2.$$

Exercise 133

For each eigenvalue λ of

$$A = \begin{bmatrix} -4 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix},$$

determine the geometric multiplicity g_λ. [**Hint**: See Exercise 132.]

We now consider a couple of examples that put together everything we've covered so far.

Example 8.3.9 Find the eigenvalues and a basis for each eigenspace of A where $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

Solution: We begin by computing the characteristic polynomial of A, using elementary row operations to aid in our computations. We have

$$C_A(\lambda) = \begin{vmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} \underset{R_3 - R_2}{\overset{R_1 + \lambda R_2}{=}} \begin{vmatrix} 0 & 1-\lambda^2 & 1+\lambda \\ 1 & -\lambda & 1 \\ 0 & 1+\lambda & -\lambda - 1 \end{vmatrix}.$$

Performing a cofactor expansion along the first column and factoring entries as needed leads to

$$\begin{aligned} &= (-1)\begin{vmatrix} (1+\lambda)(1-\lambda) & 1+\lambda \\ 1+\lambda & -(1+\lambda) \end{vmatrix} = (-1)(1+\lambda)^2 \begin{vmatrix} 1-\lambda & 1 \\ 1 & -1 \end{vmatrix} \\ &= (-1)(\lambda+1)^2\big((1-\lambda)(-1) - 1\big) = -(\lambda+1)^2(\lambda - 2). \end{aligned}$$

The eigenvalues of A are thus $\lambda_1 = -1$ with $a_{\lambda_1} = 2$ and $\lambda_2 = 2$ with $a_{\lambda_2} = 1$. For $\lambda_1 = -1$, we solve $(A+I)\vec{x} = \vec{0}$. We have

$$A + I = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \underset{\substack{R_2 - R_1 \\ R_3 - R_1}}{\longrightarrow} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus

$$\vec{x} = \begin{bmatrix} -s-t \\ s \\ t \end{bmatrix} = s\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad s,t \in \mathbb{R}$$

so

$$B_1 = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_1}(A)$ and $g_{\lambda_1} = \dim(E_{\lambda_1}(A)) = 2$. For $\lambda_2 = 2$, we solve $(A - 2I)\vec{x} = \vec{0}$. Since

$$\begin{aligned} A - 2I = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} &\underset{R_3 - R_2}{\overset{R_1 + 2R_2}{\longrightarrow}} \begin{bmatrix} 0 & -3 & 3 \\ 1 & -2 & 1 \\ 0 & 3 & -3 \end{bmatrix} \underset{R_3 + R_1}{\longrightarrow} \begin{bmatrix} 0 & -3 & 3 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \overset{-\frac{1}{3}R_1}{\longrightarrow} \\ &\begin{bmatrix} 0 & 1 & -1 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \underset{R_2 + 2R_1}{\longrightarrow} \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \overset{R_1 \leftrightarrow R_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \end{aligned}$$

we see that

$$\vec{x} = \begin{bmatrix} t \\ t \\ t \end{bmatrix} = t\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

so

$$B_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

a basis for $E_{\lambda_2}(A)$ and $g_{\lambda_2} = \dim(E_{\lambda_2}(A)) = 1$.

In Example 8.3.9, we performed elementary row operations while computing $C_A(\vec{x})$ in an effort to simplify taking the determinant. This isn't necessary as we could have immediately performed a cofactor expansion along, say, the first row of $A-\lambda I$, but it will take a bit more work to factor $C_A(\vec{x})$ in this case.

Example 8.3.10 Let $A=\begin{bmatrix}-3/5 & 4/5\\ 4/5 & 3/5\end{bmatrix}$. Find the eigenvalues of A and for each eigenvalue, find a basis for the corresponding eigenspace and state its dimension.

Solution: We first compute the characteristic polynomial. We have

$$\begin{aligned}
C_A(\lambda)=\begin{vmatrix}(-3/5)-\lambda & 4/5\\ 4/5 & (3/5)-\lambda\end{vmatrix} &= \begin{vmatrix}(-3-5\lambda)/5 & 4/5\\ 4/5 & (3-5\lambda)/5\end{vmatrix}\\
&= \frac{1}{25}\begin{vmatrix}-3-5\lambda & 4\\ 4 & 3-5\lambda\end{vmatrix}\\
&= \frac{1}{25}\big((-3-5\lambda)(3-5\lambda)-16\big)\\
&= \frac{1}{25}\big(-9+25\lambda^2-16\big)\\
&= \frac{1}{25}\big(25\lambda^2-25\big)\\
&= (\lambda+1)(\lambda-1)
\end{aligned}$$

from which we see that $\lambda_1=1$ with $a_{\lambda_1}=1$ and $\lambda_2=-1$ with $a_{\lambda_2}=1$ are the eigenvalues of A. For $\lambda_1=1$, we solve $(A-I)\vec{x}=\vec{0}$. Since

$$A-I=\begin{bmatrix}-8/5 & 4/5\\ 4/5 & -2/5\end{bmatrix} \xrightarrow{-\frac{5}{8}R_1} \begin{bmatrix}1 & -1/2\\ 4/5 & -2/5\end{bmatrix} \xrightarrow[R_2-\frac{4}{5}R_1]{} \begin{bmatrix}1 & -1/2\\ 0 & 0\end{bmatrix},$$

we see that

$$\vec{x}=\begin{bmatrix}t/2\\ t\end{bmatrix}=t\begin{bmatrix}1/2\\ 1\end{bmatrix},\quad t\in\mathbb{R}$$

so that

$$B_1=\left\{\begin{bmatrix}1/2\\ 1\end{bmatrix}\right\}$$

is a basis for $E_{\lambda_1}(A)$ and $g_{\lambda_1}=\dim(E_{\lambda_1}(A))=1$. For $\lambda_2=-1$, we solve $(A+I)\vec{x}=\vec{0}$. Since

$$A+I=\begin{bmatrix}2/5 & 4/5\\ 4/5 & 8/5\end{bmatrix} \xrightarrow{\frac{5}{2}R_1} \begin{bmatrix}1 & 2\\ 4/5 & 8/5\end{bmatrix} \xrightarrow[R_2-\frac{4}{5}R_1]{} \begin{bmatrix}1 & 2\\ 0 & 0\end{bmatrix}.$$

we have that

$$\vec{x}=\begin{bmatrix}-2t\\ t\end{bmatrix}=t\begin{bmatrix}-2\\ 1\end{bmatrix},\quad t\in\mathbb{R}$$

so that

$$B_2=\left\{\begin{bmatrix}-2\\ 1\end{bmatrix}\right\}$$

is a basis for $E_{\lambda_2}(A)$ and $g_{\lambda_2}=\dim(E_{\lambda_2}(A))=1$.

In Example 8.3.10, we used Theorem 6.3.2 to factor $\frac{1}{5}$ out of $A-\lambda I$ when computing

$C_A(\lambda)$. This is not a necessary step, but it does allow us put all fractions "out front" while computing the characteristic polynomial.

The matrix

$$A = \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}$$

in Example 8.3.10 is the standard matrix for the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which was discussed in Subsection 8.1.1. Recall that this transformation T reflects vectors through the line L containing the origin with direction vector $\vec{d} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$. We see that the eigenspace $E_{\lambda_1}(A)$ is the line L. The eigenspace $E_{\lambda_2}(A)$ is the line L' through the origin with direction vector $\vec{n} = \left[\begin{smallmatrix} -2 \\ 1 \end{smallmatrix}\right]$, which is perpendicular to L. Thus, for $\vec{x} \in E_{\lambda_1}(A) = L$, we have that $T(\vec{x}) = A\vec{x} = \vec{x}$ and for $\vec{x} \in E_{\lambda_2}(A) = L'$, we have that $T(\vec{x}) = A\vec{x} = -\vec{x}$. Thus, Example 8.3.10 confirms what we observed in Subsection 8.1.1.

Our examples thus far may lead one to believe that $a_\lambda = g_\lambda$ for every eigenvalue λ of a matrix $A \in M_{n\times n}(\mathbb{R})$. The next example shows that this is not the case.[2]

Example 8.3.11

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Find the eigenvalues of A, and for each eigenvalue, find a basis for the corresponding eigenspace.

Solution: Since A is upper triangular, Theorem 8.2.9 gives that $\lambda = 1$ is the only eigenvalue of A and we see that $a_{\lambda_1} = 2$. Thus we solve $(A - I)\vec{x} = \vec{0}$. We have

$$A - I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

which gives

$$\vec{x} = \begin{bmatrix} t \\ 0 \end{bmatrix} = t\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}.$$

It follows that

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

is a basis for $E_\lambda(A)$ and $g_\lambda = \dim(E_\lambda(A)) = 1$.

Example 8.3.11 shows us that it is possible for $g_\lambda < a_\lambda$. The next theorem guarantees that the g_λ cannot exceed a_λ and will be useful in the next section.

Theorem 8.3.12

Let $A \in M_{n\times n}(\mathbb{R})$. For any eigenvalue λ of A,

$$1 \le g_\lambda \le a_\lambda \le n.$$

Theorem 8.3.12 will play an important role in Section 8.4. The proofs of the statements $1 \le g_\lambda$ and $a_\lambda \le n$ are left as exercises at the end of this section. The proof that $g_\lambda \le a_\lambda$ is unfortunately beyond the scope of this course.

[2] If you've been keeping up with your exercises, then Exercise 132 and Exercise 133 will have already convinced you that it's possible for $a_\lambda \ne g_\lambda$.

Section 8.3 Problems

8.3.1. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$.

(a) Find the eigenvalues of A and state their algebraic multiplicities.

(b) For each eigenvalue of A, compute its geometric multiplicity by finding a basis for the corresponding eigenspace.

8.3.2. Let $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix}$.

(a) Find the eigenvalues of A and state their algebraic multiplicities.

(b) For each eigenvalue of A, compute its geometric multiplicity by finding a basis for the corresponding eigenspace.

8.3.3. Let $A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

(a) Find the eigenvalues of A and state their algebraic multiplicities.

(b) For each eigenvalue of A, compute its geometric multiplicity by finding a basis for the corresponding eigenspace.

8.3.4. Find the eigenvalues of

$$A = \begin{bmatrix} 5/6 & -1/3 & -1/6 \\ -1/3 & 1/3 & -1/3 \\ -1/6 & -1/3 & 5/6 \end{bmatrix}.$$

For each eigenvalue, find a basis for $E_\lambda(A)$. Note that A is the standard matrix of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ that projects $\vec{x} \in \mathbb{R}^3$ onto the plane P with scalar equation $x_1 + 2x_2 + x_3 = 0$ (see Subsection 8.1.2). Are the results obtained here consistent with what we observed in Subsection 8.1.2? Note that computing $C_A(\lambda)$ will be tedious - you should arrive at $C_A(\lambda) = -\lambda(\lambda - 1)^2$.

8.3.5. Let $A \in M_{n\times n}(\mathbb{R})$ and let $\lambda \in \mathbb{R}$ be an eigenvalue of A. Prove that

(a) $a_\lambda \leq n$,

(b) $g_\lambda \geq 1$.

8.4 Diagonalization

This section is concerned with using our knowledge of eigenvalues and eigenvectors to represent certain matrices $A \in M_{n\times n}(\mathbb{R})$ in terms of diagonal matrices $D \in M_{n\times n}(\mathbb{R})$. The results we derive here are useful in many areas of science and engineering, such as machine learning and quantum mechanics, in addition to being useful in later mathematics courses where, for example, students will use these results to solve recurrence relations and to compute matrix exponentials to solve linear systems of differential equations.

Before moving forward, we briefly discuss diagonal matrices. Recall that diagonal matrices were defined in Definition 6.2.7 as square matrices that were both upper and lower triangular. The next definition equivalently defines diagonal matrices explicitly in terms of their entries.

Definition 8.4.1
Diagonal Matrix

A matrix $D = [d_{ij}] \in M_{n\times n}(\mathbb{R})$ is a **diagonal matrix** if $d_{ij} = 0$ for all $i \neq j$. In this case, we may write $D = \operatorname{diag}(d_{11}, \ldots, d_{nn})$.

It is important to note that Definition 8.4.1 places no conditions on the values of the main diagonal entries $d_{11}, \ldots, d_{nn}$ of D. It simply states that any entry *not* on the main diagonal of D must be zero.

Example 8.4.2

The matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 101 & 0 & 0 & 0 \\ 0 & 115 & 0 & 0 \\ 0 & 0 & 116 & 0 \\ 0 & 0 & 0 & 117 \end{bmatrix}$$

are diagonal matrices.

The next theorem shows that diagonal matrices behave very well with respect to the operations of matrix addition, scalar multiplication and matrix multiplication.

Theorem 8.4.3

Let $D, F \in M_{n\times n}(\mathbb{R})$ be diagonal matrices with

$$D = \operatorname{diag}(d_{11}, \ldots, d_{nn}) \quad \text{and} \quad F = \operatorname{diag}(f_{11}, \ldots, f_{nn}),$$

and let $c \in \mathbb{R}$. Then $D + F$, cD and DF are diagonal matrices. In particular,

$$\begin{aligned} D + F &= \operatorname{diag}(d_{11} + f_{11}, \ldots, d_{nn} + f_{nn}), \\ cD &= \operatorname{diag}(cd_{11}, \ldots, cd_{nn}), \\ DF &= \operatorname{diag}(d_{11}f_{11}, \ldots, d_{nn}f_{nn}). \end{aligned}$$

Moreover, $DF = FD$, that is, diagonal matrices commute under matrix multiplication.

The result concerning the product of diagonal matrices in Theorem 8.4.3 can be extended to more than two matrices. For diagonal matrices $D_1, \ldots, D_k \in M_{n\times n}(\mathbb{R})$, we have that the product $D_1 \cdots D_k$ is a diagonal matrix with[3]

$$D_1 \cdots D_k = \operatorname{diag}((d_1)_{11} \cdots (d_k)_{11}, \ldots, (d_1)_{nn} \cdots (d_k)_{nn}).$$

[3]For matrices denoted with a subscript, say D_ℓ, we denote the (i, j)-entry of D_ℓ by $(d_\ell)ij$.

In particular, if $D_1 = \cdots = D_k = D$, then we obtain

$$D^k = \mathrm{diag}(d_{11}^k, \ldots, d_{nn}^k),$$

a result that will be useful in Section 8.5.

We now make the following important definition.

Definition 8.4.4
Diagonalizable Matrix

A matrix $A \in M_{n\times n}(\mathbb{R})$ is **diagonalizable** if there exists an invertible matrix $P \in M_{n\times n}(\mathbb{R})$ and a diagonal matrix $D \in M_{n\times n}(\mathbb{R})$ so that $P^{-1}AP = D$. In this case, we say that P diagonalizes A to D.

It is important to note that $P^{-1}AP = D$ does not imply that $A = D$ in general. This is because matrix multiplication is not commutative, so we cannot cancel P and P^{-1} in the expression $P^{-1}AP$.

Given a matrix $A \in M_{n\times n}(\mathbb{R})$, we now consider how to determine if we can find an invertible $P \in M_{n\times n}(\mathbb{R})$ and a diagonal $D \in M_{n\times n}(\mathbb{R})$ so that $P^{-1}AP = D$, and if so, how to construct such matrices P and D. As alluded to at the start of this section, eigenvalues and eigenvectors will play a significant role. We will need the following two results.

Theorem 8.4.5

Let $A \in M_{n\times n}(\mathbb{R})$ and assume that $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ are the distinct eigenvalues A. Then the algebraic multiplicities of $\lambda_1, \ldots, \lambda_k$ satisfy

$$a_{\lambda_1} + \cdots + a_{\lambda_k} = n$$

while the geometric multiplicities satisfy the inequalities

$$k \le g_{\lambda_1} + \cdots + g_{\lambda_k} \le n$$

with $g_{\lambda_1} + \cdots + g_{\lambda_k} = n$ if and only if $g_{\lambda_i} = a_{\lambda_i}$ for $i = 1, \ldots, k$.

Proof: Since $A \in M_{n\times n}(\mathbb{R})$, we see that $C_A(\lambda)$ is a real polynomial of degree $n \ge 1$. Since every real polynomial is a complex polynomial, Theorem 7.4.8 states that $C_A(\lambda)$ has exactly n roots counting multiplicities, and hence

$$a_{\lambda_1} + \cdots + a_{\lambda_k} = n.$$

It follows from Theorem 8.3.12 that $1 \le g_{\lambda_i} \le a_{\lambda_i}$ for $i = 1, \ldots, k$. Summing over i gives

$$k \le g_{\lambda_1} + \cdots + g_{\lambda_k} \le a_{\lambda_1} + \cdots + a_{\lambda_k} = n.$$

Since Theorem 8.3.12 guarantees that the geometric multiplicity cannot exceed the algebraic multiplicity for any eigenvalue λ of A, we have $g_{\lambda_1} + \cdots + g_{\lambda_k} = n$ if and only if $g_{\lambda_i} = a_{\lambda_i}$ for $i = 1, \ldots, k$. $\square$

The next theorem requires knowledge of the *union* of sets. Take a look at Definition A.1.6 if the union of sets is unfamiliar.

Theorem 8.4.6 Let $A \in M_{n\times n}(\mathbb{R})$ and assume that $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ are the distinct eigenvalues A. For each $i = 1, \ldots, k$, let B_i be a basis for the corresponding eigenspace $E_{\lambda_i}(A)$. Then

$$B = B_1 \cup B_2 \cup \cdots \cup B_k$$

is linearly independent. In particular, B is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A if and only if $g_{\lambda_i} = a_{\lambda_i}$ for $i = 1, \ldots, k$.

Example 8.4.7 We saw in Example 8.3.4 that the matrix

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

has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$, and that

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad B_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

are bases for $E_{\lambda_1}(A)$ and $E_{\lambda_2}(A)$, respectively. If we define

$$B = B_1 \cup B_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\},$$

then Theorem 8.4.6 guarantees that B is linearly independent. In fact, since B has 3 vectors, it will follow from the the Matrix Invertibility Criteria Revisited that B a basis for $\mathbb{R}^3$, which by construction consists of eigenvectors of A.

Exercise 134 Verify that the set B from Example 8.4.7 is a basis for $\mathbb{R}^3$ without using Theorem 8.4.6.

Proof of Theorem 8.4.6: We omit the proof that $B = B_1 \cup B_2 \cup \cdots \cup B_k$ is linearly independent and only prove that B is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A if and only if $g_{\lambda_i} = a_{\lambda_i}$ for $i = 1, \ldots, k$.

Since B_i is a basis for $E_{\lambda_i}(A)$, we have that B_i consists of eigenvectors of A corresponding to λ_i. Thus, B consists of eigenvectors of A. Also, since $\dim(E_{\lambda_i}(A)) = g_{\lambda_i}$, we have that B_i contains g_{λ_i} vectors, and so B contains $g_{\lambda_1} + \cdots + g_{\lambda_k}$ vectors. Thus, by Theorem 8.4.5, B contains at least k vectors and at most n vectors.

Since $\dim(\mathbb{R}^n) = n$, every basis for $\mathbb{R}^n$ must contain n vectors. Thus, by Theorem 8.4.5, B is a basis for $\mathbb{R}^n$ (consisting of eigenvectors of A) if and only if $g_{\lambda_i} = a_{\lambda_i}$ for $i = 1, \ldots k$. $\square$

We now present the main result of this section.

Theorem 8.4.8 **(Diagonalization Theorem)**

Let $A \in M_{n\times n}(\mathbb{R})$ and assume that none of the eigenvalues of A are non-real. Then A is diagonalizable if and only if there exists a basis for $\mathbb{R}^n$ consisting of eigenvectors of A.

Proof: We first assume that A is diagonalizable. Then there exists an invertible matrix $P = \begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \end{bmatrix} \in M_{n\times n}(\mathbb{R})$ and a diagonal matrix $D = \operatorname{diag}(\mu_1, \ldots, \mu_n) \in M_{n\times n}(\mathbb{R})$ such that $P^{-1}AP = D$, that is, such that $AP = PD$. Thus

$$\begin{aligned} A\begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \end{bmatrix} &= P\begin{bmatrix} \mu_1\vec{e}_1 & \cdots & \mu_n\vec{e}_n \end{bmatrix} \\ \begin{bmatrix} A\vec{x}_1 & \cdots & A\vec{x}_n \end{bmatrix} &= \begin{bmatrix} \mu_1 P\vec{e}_1 & \cdots & \mu_n P\vec{e}_n \end{bmatrix} \\ \begin{bmatrix} A\vec{x}_1 & \cdots & A\vec{x}_n \end{bmatrix} &= \begin{bmatrix} \mu_1\vec{x}_1 & \cdots & \mu_n\vec{x}_n \end{bmatrix}. \end{aligned}$$

By equating columns, we see that

$$A\vec{x}_i = \mu_i\vec{x}_i \tag{8.7}$$

for $i = 1, \ldots, n$, and since $P = \begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \end{bmatrix}$ is invertible, it follows from the Matrix Invertibility Criteria Revisited that the set $\{\vec{x}_1, \ldots, \vec{x}_n\}$ is a basis for $\mathbb{R}^n$. This implies that that $\vec{x}_i \neq \vec{0}$ for $i = 1, \ldots, n$ and it follows from (8.7) that μ_i is an eigenvalue of A and that $\vec{x}_i$ is a corresponding eigenvector. Thus $\{\vec{x}_1, \ldots, \vec{x}_n\}$ is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A.

We now assume that there is a basis $\{\vec{x}_1, \ldots, \vec{x}_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors of A. Let $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ be the distinct eigenvalues of A. Then for each $i = 1, \ldots n$, $A\vec{x}_i = \mu_i\vec{x}_i$ where $\mu_i = \lambda_j$ for some $j = 1, \ldots, k$. It follows from the Matrix Invertibility Criteria Revisited that $P = \begin{bmatrix} \vec{x}_1 & \cdots & \vec{x}_n \end{bmatrix}$ is invertible and thus

$$\begin{aligned} P^{-1}AP &= P^{-1}\begin{bmatrix} A\vec{x}_1 & \cdots & A\vec{x}_n \end{bmatrix} \\ &= P^{-1}\begin{bmatrix} \mu_1\vec{x}_1 & \cdots & \mu_n\vec{x}_n \end{bmatrix} \\ &= P^{-1}\begin{bmatrix} \mu_1 P\vec{e}_1 & \cdots & \mu_n P\vec{e}_n \end{bmatrix} \\ &= P^{-1}P\begin{bmatrix} \mu_1\vec{e}_1 & \cdots & \mu_n\vec{e}_n \end{bmatrix} \\ &= \operatorname{diag}(\mu_1, \ldots, \mu_n) \end{aligned}$$

which shows that A is diagonalizable. □

The proof of the Diagonalization Theorem is a *constructive* proof, that is, given a diagonalizable matrix A, it tells us exactly how to construct the invertible matrix P and the diagonal matrix D so that $P^{-1}AP = D$. Given that A is diagonalizable, the jth column of P will contain the jth vector from the basis of eigenvectors, and the jth column of the diagonal matrix D will contain the corresponding eigenvalue in the $(j, j)-$entry.

The following are consequences of the Diagonalization Theorem.

Corollary 8.4.9 Let $A \in M_{n\times n}(\mathbb{R})$ be such that all of the eigenvalues of A are real. Then A is diagonalizable if and only if $a_\lambda = g_\lambda$ for every eigenvalue λ of A.

Corollary 8.4.10 Let $A \in M_{n\times n}(\mathbb{R})$ be such that all of the eigenvalues of A are real. If A has n distinct eigenvalues, then A is diagonalizable.

The following algorithm summarizes the steps needed to determine if a matrix A with real eigenvalues is diagonalizable.

ALGORITHM (Diagonalization)

Let $A \in M_{n\times n}(\mathbb{R})$ and assume that none of the eigenvalues of A are non-real. To diagonalize A, perform the following steps.

- **Step 1**: Factor $C_A(\lambda)$ to determine the eigenvalues $\lambda_1, \dots, \lambda_k$ of A and determine $a_{\lambda_1}, \dots, a_{\lambda_k}$.
- **Step 2**: For $i = 1, \dots, k$, solve the homogeneous system $(A - \lambda_i I)\vec{x} = \vec{0}$ to determine a basis B_i for $E_{\lambda_i}(A)$, and determine g_{λ_i}.
- **Step 3**: Decide if A is diagonalizable:
  - If $g_{\lambda_i} < a_{\lambda_i}$ for some $i = 1, \dots, k$, then A is not diagonalizable - do not continue with the algorithm.
  - if $g_{\lambda_i} = a_{\lambda_i}$ for every $i = 1, \dots, k$, then A is diagonalizable.
- **Step 4**: Construct an $n \times n$ matrix P whose columns are the n vectors in $B = B_1 \cup \cdots \cup B_k$, in any order.
- **Step 5**: Construct a diagonal matrix D whose jth column is $\lambda \vec{e}_j$, where λ is the eigenvalue of A corresponding to the eigenvector of A in the jth column of P.

When asked to diagonalize a matrix $A \in M_{n\times n}(\mathbb{R})$, we need only find an invertible matrix P and a diagonal matrix D so that $P^{-1}AP = D$. We do not need to compute P^{-1} in order to verify that $P^{-1}AP = D$ as this is guaranteed by the Diagonalization Theorem. However, it is a good idea to do this anyway in order to verify that our work is correct.

Example 8.4.11 Recall the matrix $A = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}$ from Example 8.3.3. We computed $C_A(\lambda) = (\lambda+1)(\lambda-1)$ so that the eigenvalues of A are $\lambda_1 = -1$ and $\lambda_2 = 7$, from which it follows that $a_{\lambda_1} = 1 = a_{\lambda_2}$. We found that

$$B_1 = \left\{ \begin{bmatrix} -3/2 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad B_2 = \left\{ \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} \right\}$$

are bases for $E_{\lambda_1}(A)$ and $E_{\lambda_2}(A)$, respectively so that $g_{\lambda_1} = 1 = g_{\lambda_2}$. Since $a_{\lambda_1} = g_{\lambda_1}$ and $a_{\lambda_2} = g_{\lambda_2}$, A is diagonalizable by Corollary 8.4.9. Thus we let

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

so that

$$P^{-1}AP = \begin{bmatrix} -1 & 0 \\ 0 & 7 \end{bmatrix} = D.$$

We make a few remarks about Example 8.4.11. First, notice that $A \in M_{2\times 2}(\mathbb{R})$ has 2 distinct eigenvalues. Thus we could have used Corollary 8.4.10 to conclude that A is diagonalizable before we even computed bases for the corresponding eigenspaces.

Secondly, note that P and D are not unique. We could have instead chosen

$$P = \begin{bmatrix} 1/2 & -3/2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 7 & 0 \\ 0 & -1 \end{bmatrix}.$$

In fact, for each eigenspace $E_\lambda(A)$, we can select *any* basis, and we can order the resulting columns of P in any order we like so long as the eigenvalues in each column of D correspond to the eigenvector of A in the corresponding column of P.

Lastly, we can (and should) check our work. It's not too difficult to compute

$$P^{-1} = \begin{bmatrix} -1/2 & 1/4 \\ 1/2 & 3/4 \end{bmatrix}$$

and then verify that

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} -1/2 & 1/4 \\ 1/2 & 3/4 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} -3/2 & 1/2 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1/2 & -1/4 \\ 7/2 & 21/4 \end{bmatrix} \begin{bmatrix} -3/2 & 1/2 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & 7 \end{bmatrix} \\ &= D. \end{aligned}$$

Example 8.4.12 Diagonalize the matrix $A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Solution: From Example 8.3.4, we have that the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 2$ with $a_{\lambda_1} = 1$ and $a_{\lambda_2} = 2$. We also have that

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad B_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

are bases for $E_{\lambda_1}(A)$ and $E_{\lambda_2}(A)$ respectively, so $g_{\lambda_1} = 1$ and $g_{\lambda_2} = 2$. Since $a_{\lambda_1} = g_{\lambda_1}$ and $a_{\lambda_2} = g_{\lambda_2}$, we see that A is diagonalizable. We let

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

so that

$$P^{-1}AP = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = D.$$

We follow Example 8.4.12 with a few remarks. First, notice that $A \in M_{3\times 3}(\mathbb{R})$, but that A only has 2 distinct eigenvalues. Thus, we cannot use Corollary 8.4.10 to conclude that A is diagonalizable as we could have in Example 8.4.11. We must compute a basis for each eigenspace of A and ensure that $g_\lambda = a_\lambda$ for each of the two eigenvalues before we may conclude that A is diagonalizable.

Second, we again see that the matrix P is not unique. We could have used

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{with} \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

for example.

Finally, it's again a good idea to check $P^{-1}AP = D$ even though it's a bit more work to compute P^{-1} for a 3×3 matrix.

Exercise 135 Diagonalize the matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$. [**Hint**: See Example 8.3.9.]

Of course, not every matrix $A \in M_{n\times n}(\mathbb{R})$ will be diagonalizable, as the next example illustrates.

Example 8.4.13 Recall from Example 8.3.11 that

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

has eigenvalues $\lambda_1 = 1$ with $a_{\lambda_1} = 2$. However, a basis for $E_{\lambda_1}(A)$ is

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

so $g_{\lambda_1} = 1 \neq 2 = a_{\lambda_1}$. It follows from Corollary 8.4.9 that A is not diagonalizable.

Finally, we can recover a diagonalizable matrix A given only its eigenvalues and bases for the corresponding eigenspaces. Notice that since A is diagonalizable, we can write $P^{-1}AP = D$ where P and D are constructed as in our Diagonalization Algorithm. Rearranging then gives $A = PDP^{-1}$, which we use in the next example.

Example 8.4.14 Let $A \in M_{3\times 3}(\mathbb{R})$ have two distinct eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -1$ and suppose

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right\} \quad \text{and} \quad B_2 = \left\{ \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \right\}$$

are bases for $E_{\lambda_1}(A)$ and $E_{\lambda_2}(A)$, respectively. Determine the matrix A.

Solution: The set

$$B = B_1 \cup B_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \right\}$$

is linearly independent by Theorem 8.4.6. Since B contains 3 vectors, it follows from the Matrix Invertibility Criteria Revisited that B is a basis for $\mathbb{R}^3$. Since B consists of eigenvectors of A, we have that A is diagonalizable by the Diagonalization Theorem. Let

$$P = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

We compute P^{-1} using the Matrix Inversion Algorithm.

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 3 & 0 & 1 & 0 \\ 1 & 2 & 4 & 0 & 0 & 1 \end{array}\right] \underset{\substack{R_2 - R_1 \\ R_3 - R_1}}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 & 1 \end{array}\right] \overset{R_1 - R_2}{\underset{R_3 - R_2}{\longrightarrow}}$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 2 & -1 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right] \overset{\substack{R_1 - R_3 \\ R_2 - R_3}}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & -1 & 2 & -1 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right],$$

so

$$P^{-1} = \begin{bmatrix} 2 & 0 & -1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

We compute

$$\begin{aligned} A = PDP^{-1} &= \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & -1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 2 & -2 \\ 2 & 4 & -3 \\ 2 & 4 & -4 \end{bmatrix} \begin{bmatrix} 2 & 0 & -1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 6 & -6 \\ 0 & 11 & -9 \\ 0 & 12 & -10 \end{bmatrix}. \end{aligned}$$

Exercise 136 Let $A \in M_{2\times 2}(\mathbb{R})$ with eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 6$. Suppose

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\} \quad \text{and} \quad B_2 = \left\{ \begin{bmatrix} -2 \\ 2 \end{bmatrix} \right\}$$

are bases for $E_{\lambda_1}(A)$ and $E_{\lambda_2}(A)$, respectively. Determine the matrix A.

Section 8.4 Problems

8.4.1. Diagonalize the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$, if possible. [**Hint**: See Problem 8.3.1.]

8.4.2. Diagonalize the matrix $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix}$, if possible. [**Hint**: See Problem 8.3.2.]

8.4.3. Diagonalize the matrix $A = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$, if possible. [**Hint**: See Problem 8.3.3.]

8.4.4. Diagonalize the matrix $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix}$, if possible. [**Hint**: See Problem 8.2.(a).]

8.4.5. A matrix $A \in M_{3\times 3}(\mathbb{R})$ has three distinct eigenvalues $\lambda_1 = 2$, $\lambda_2 = -2$ and $\lambda_3 = 0$. Determine the matrix A given that

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}, \quad B_2 = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\} \quad \text{and} \quad B_3 = \left\{ \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}$$

are bases for the eigenspaces $E_{\lambda_1}(A)$, $E_{\lambda_2}(A)$ and $E_{\lambda_3}(A)$, respectively.

8.4.6. Prove Theorem 8.4.3.

8.4.7. Prove Corollary 8.4.9.

8.4.8. Prove Corollary 8.4.10.

8.5 Powers of Matrices

In this section, we will see how diagonalizing a matrix $A \in M_{n\times n}(\mathbb{R})$ can help us compute A^k for any positive integer k. This is useful in many areas, for example, in stochastic processes where we predict the probability of a sequence of events occurring given that we know the outcome of the most recent event.

Suppose $A \in M_{n\times n}(\mathbb{R})$ is an diagonalizable matrix. Then $P^{-1}AP = D$ for some invertible $P \in M_{n\times n}(\mathbb{R})$ and some diagonal matrix $D \in M_{n\times n}(\mathbb{R})$. Rearranging gives $A = PDP^{-1}$ and we can compute

$$\begin{aligned}
A^2 &= AA = PDP^{-1}PDP^{-1} = PDIDP^{-1} = PD^2P^{-1}, \\
A^3 &= A^2A = PD^2P^{-1}PDP^{-1} = PD^2IDP^{-1} = PD^3P^{-1}, \\
&\vdots
\end{aligned}$$

As we continue this process, we will see that $A^k = PD^kP^{-1}$ for any positive integer k. Although computing powers of an $n \times n$ matrix by inspection can be difficult, if not impossible, the discussion immediately following Theorem 8.4.3 shows that computing a positive integer power of a diagonal matrix is quite easy. Recall that if

$$D = \operatorname{diag}(d_{11}, \dots, d_{nn}),$$

then

$$D^k = \operatorname{diag}(d_{11}^k, \dots, d_{nn}^k)$$

for any positive integer k.

Example 8.5.1

Let $A = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}$. Find a formula for A^k.

Solution: From Example 8.4.11, A is diagonalizable with

$$P = \frac{1}{2}\begin{bmatrix} -3 & 1 \\ 2 & 2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} -1 & 0 \\ 0 & 7 \end{bmatrix}.$$

We leave it as an exercise to verify that

$$P^{-1} = \frac{1}{4}\begin{bmatrix} -2 & 1 \\ 2 & 3 \end{bmatrix}.$$

Thus

$$\begin{aligned}
A^k &= PD^kP^{-1} \\
&= \frac{1}{2}\left(\frac{1}{4}\right)\begin{bmatrix} -3 & 1 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} (-1)^k & 0 \\ 0 & 7^k \end{bmatrix}\begin{bmatrix} -2 & 1 \\ 2 & 3 \end{bmatrix} \\
&= \frac{1}{8}\begin{bmatrix} 3(-1)^{k+1} & 7^k \\ 2(-1)^k & 2(7)^k \end{bmatrix}\begin{bmatrix} -2 & 1 \\ 2 & 3 \end{bmatrix} \\
&= \frac{1}{8}\begin{bmatrix} 6(-1)^{k+2} + 2(7)^k & 3(-1)^{k+1} + 3(7)^k \\ 4(-1)^{k+1} + 4(7)^k & 2(-1)^k + 6(7)^k \end{bmatrix}.
\end{aligned}$$

Note that we can verify our work is reasonable by taking $k = 1$ and ensuring we get A.

$$A^1 = \frac{1}{8}\begin{bmatrix} 6(-1)^{1+2} + 2(7)^1 & 3(-1)^{1+1} + 3(7)^1 \\ 4(-1)^{1+1} + 4(7)^1 & 2(-1)^1 + 6(7)^1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} = A.$$

We can use our formula for A^k to compute say, A^5:

$$A^5 = \begin{bmatrix} 4201 & 6303 \\ 8404 & 12605 \end{bmatrix}.$$

Example 8.5.2

Let $A = \begin{bmatrix} 3 & -4 \\ -2 & 1 \end{bmatrix}$. Find a formula for A^k.

Solution: We first compute the eigenvalues of A.

$$C_A(\lambda) = \begin{vmatrix} 3-\lambda & -4 \\ -2 & 1-\lambda \end{vmatrix} = (3-\lambda)(1-\lambda) - 8 = \lambda^2 - 4\lambda + 3 - 8 = \lambda^2 - 4\lambda - 5 = (\lambda - 5)(\lambda + 1)$$

so $\lambda_1 = -1$ and $\lambda_2 = 5$ are the eigenvalues of A. Since A is a 2×2 matrix with 2 distinct eigenvalues, A is diagonalizable by Corollary 8.4.10. For $\lambda_1 = -1$, we solve $(A + I)\vec{x} = \vec{0}$. Since

$$A + I = \begin{bmatrix} 4 & -4 \\ -2 & 2 \end{bmatrix} \underset{R_2 + \frac{1}{2}R_1}{\longrightarrow} \begin{bmatrix} 4 & -4 \\ 0 & 0 \end{bmatrix} \overset{\frac{1}{4}R_1}{\longrightarrow} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

we see that

$$\vec{x} = \begin{bmatrix} t \\ t \end{bmatrix} = t\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R},$$

so

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_1}(A)$. For $\lambda_2 = 5$, we solve $(A - 5I)\vec{x} = \vec{0}$. Since

$$A - 5I = \begin{bmatrix} -2 & -4 \\ -2 & -4 \end{bmatrix} \underset{R_2 - R_1}{\longrightarrow} \begin{bmatrix} -2 & -4 \\ 0 & 0 \end{bmatrix} \overset{-\frac{1}{2}R_1}{\longrightarrow} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix},$$

we have that

$$\vec{x} = \begin{bmatrix} -2t \\ t \end{bmatrix} = t\begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R},$$

so

$$B_2 = \left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_2}(A)$. Now, let

$$P = \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} \text{ so that } P^{-1}AP = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix} = D.$$

Then

$$P^{-1} = \frac{1}{3}\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$$

and

$$A^k = PD^kP^{-1} = \frac{1}{3}\begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} (-1)^k & 0 \\ 0 & 5^k \end{bmatrix}\left(\frac{1}{3}\right)\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$$

$$
\begin{aligned}
&= \frac{1}{3}\begin{bmatrix} (-1)^k & (-2)5^k \\ (-1)^k & 5^k \end{bmatrix}\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \\
&= \frac{1}{3}\begin{bmatrix} (-1)^k + (2)5^k & 2(-1)^k - (2)5^k \\ (-1)^k - 5^k & 2(-1)^k + 5^k \end{bmatrix}.
\end{aligned}
$$

Section 8.5 Problems

8.5.1. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$. Find a formula for A^k. [**Hint**: See Problem 8.4.1.]

8.5.2. Let $A = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Find a formula for A^k. [**Hint**: See Example 8.4.12.]

8.5.3. Let $A = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Find a formula for A^k.

Appendix A

A Brief Introduction to Sets

Sets will play an important role in linear algebra, so we need to understand the basic results concerning them. We begin with the definition of a set. Note that this definition is far from the formal definition, and can lead to contradictions if we are not careful. For our purposes, however, this definition will be sufficient.

Definition A.1.1
Set

A **set** is a collection of objects. We call the objects **elements** of the set

Example A.1.2

The following are examples of sets.

- $A = \{1, 2, 3\}$ is a set with three elements, namely 1, 2 and 3,
- $B = \{\heartsuit, f(x), \{1, 2\}, 3\}$,
- $\varnothing = \{\,\}$, the set with no elements, which is called the *empty set*.

We see that one way to describe a set is to list the elements of the set between curly braces "{" and "}". The set B shows that a set can have elements other than numbers: the elements can be functions, other sets, or other symbols. The empty set has no elements in it, and we normally prefer using $\varnothing$ over $\{\,\}$ in this case.

Given a set A, we write $x \in A$ if x is an element of A, and $x \notin A$ is x is not an element of A.

Example A.1.3

For $B = \{\heartsuit, f(x), \{1, 2\}, 3\}$, we have

$$\heartsuit \in B, \quad f(x) \in B, \quad \{1, 2\} \in B \quad \text{and} \quad 3 \in B$$

but

$$1 \notin T \quad \text{and} \quad 2 \notin B.$$

Example A.1.4

Here are a few sets that you may be familiar with.

- **Natural numbers**: $\mathbb{N} = \{1, 2, 3, \ldots\}$,
- **Integers**: $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$,
- **Rational numbers**: $\mathbb{Q} = \left\{ \frac{a}{b} \,\middle|\, a, b \in \mathbb{Z}, b \neq 0 \right\}$,
- **Real numbers**: $\mathbb{R}$ is the the set of all numbers that are either rational or irrational,
- **Complex numbers**: $\mathbb{C} = \{a + bi \,|\, a, b \in \mathbb{R}\}$,
- $\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \,\middle|\, x_1, \ldots, x_n \in \mathbb{R} \right\}$.

Note that each of these sets in Example A.1.4 contains infinitely many elements. The sets $\mathbb{N}$ and $\mathbb{Z}$ are defined by listing their elements (or rather, listing enough elements so that you "get the idea"), the set $\mathbb{R}$ is defined using words, and the sets $\mathbb{Q}$, $\mathbb{C}$ and $\mathbb{R}^n$ are defined using *set builder notation* where conditions are given that elements of the set must satisfy. For example, the set

$$\mathbb{Q} = \left\{ \frac{a}{b} \,\middle|\, a, b \in \mathbb{Z}, b \neq 0 \right\}$$

is understood to mean "$\mathbb{Q}$ is the set of all fractions of the form $\frac{a}{b}$ satisfying the conditions that a and b are integers and b is nonzero". If a fraction $\frac{a}{b}$ satisfies these conditions, then it is a rational number, otherwise it is not.

For a set A defined via set builder notation, we can determine whether an element belongs to A by seeing if it satisfies the condition

Example A.1.5

Let

$$U = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \,\middle|\, 2x_1 - x_2 + x_3 = 4 \right\}.$$

Determine whether $\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right] \in U$.

Solution: Since $2(1) - 2 + 3 = 3 \neq 4$, we have that $\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right] \notin U$.

We now define two ways that we can combine given sets to create new sets.

Definition A.1.6
Union, Intersection

Let A, B be sets. The **union** of A and B is the set

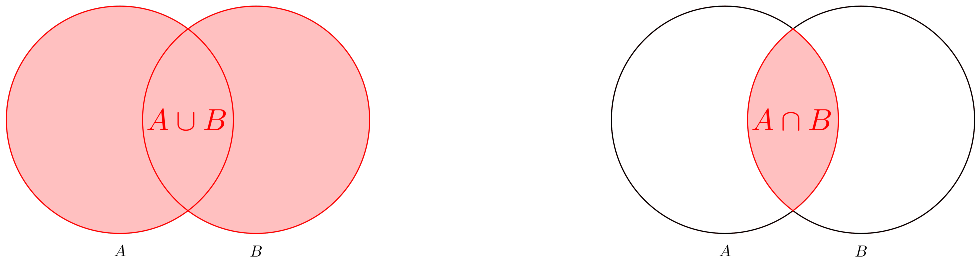
$$A \cup B = \{x \,|\, x \in A \text{ or } x \in B\}$$

and the **intersection** of A and B is the set

$$A \cap B = \{x \,|\, x \in A \text{ and } x \in B\}.$$

We think of the union of two sets A and B as the set of elements that belong to *at least one* of A or B, and we think of the intersection of two sets A and B as the set of elements that belong to *both* A and B.

We can visualize the union and intersection of two sets using *Venn Diagrams*. Although Venn Diagrams can help us visualize sets, they should never be used as part of a proof of any statement regarding sets.



(a) A Venn Diagram depicting two sets, A and B. Their union is the shaded region.

(b) A Venn Diagram depicting two sets, A and B. Their intersection is the shaded region.

Figure A.1.1: Venn Diagrams.

Example A.1.7

If $A = \{1, 2, 3, 4\}$ and $B = \{-1, 2, 4, 6, 7\}$, then

$$A \cup B = \{-1, 1, 2, 3, 4, 6, 7\}$$
$$A \cap B = \{2, 4\}$$

The notion of a union of sets and an intersection of sets is not restricted to just two sets. If $A_1, \ldots, A_k$ are sets, then

$$A_1 \cup A_2 \cup \cdots \cup A_k = \{x \,|\, x \in A_i \text{ for some } i = 1, \ldots, k\}$$
$$A_1 \cap A_2 \cap \cdots \cap A_k = \{x \,|\, x \in A_i \text{ for each } i = 1, \ldots, k\}.$$

Definition A.1.8
Subset

Let S, T be sets. We say that S is a **subset** of T (and we write $S \subseteq T$) if for every $x \in S$ we have that $x \in T$. If S is not a subset of T, then we write $S \not\subseteq T$.

Example A.1.9

Let $A = \{1, 2, 4\}$ and $B = \{1, 2, 3, 4\}$. Then $A \subseteq B$ since every element of A is an element of B, but $B \not\subseteq A$ since $3 \in B$, but $3 \notin A$.

Note that it's important to distinguish between an element of a set and a subset of a set. For example,

$$1 \in \{1, 2, 3\} \quad \text{but} \quad 1 \not\subseteq \{1, 2, 3\}$$

and

$$\{1\} \notin \{1, 2, 3\} \quad \text{but} \quad \{1\} \subseteq \{1, 2, 3\}.$$

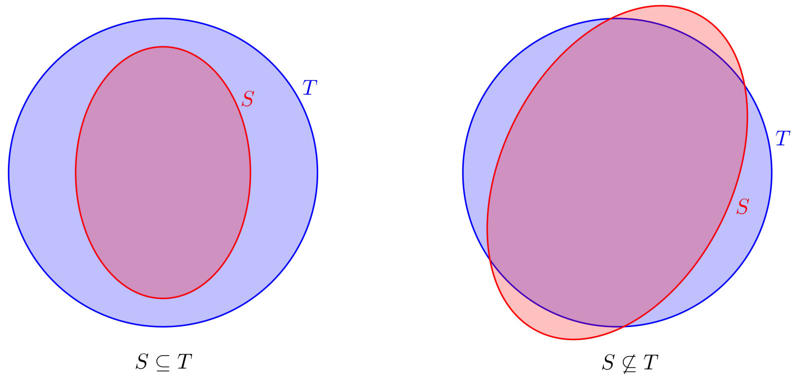


Figure A.1.2: A Venn diagram showing an instance when $S \subseteq T$ on the left, and an instance when $S \not\subseteq T$ (and also $T \not\subseteq S$) on the right.

More interestingly,

$$\{1, 2\} \in \{1, 2, \{1, 2\}\} \quad \text{and} \quad \{1, 2\} \subseteq \{1, 2, \{1, 2\}\}$$

which shows that an element of a set may also be a subset of a set. This last example can cause students to stumble, so the following may help:

$$\boldsymbol{\{1,2\}} \in \{1, 2, \boldsymbol{\{1,2\}}\} \quad \text{and} \quad \{\mathbf{1}, \mathbf{2}\} \subseteq \{\mathbf{1}, \mathbf{2}, \{1, 2\}\}.$$

Finally we mention that for any set A, we have that $\varnothing \subseteq A$. This generally seems quite strange at first. However if $\varnothing \not\subseteq A$, then there must be some element $x \in \varnothing$ such that $x \notin A$. But the empty set contains no elements, so we can never show that $\varnothing$ is not a subset of A. Thus we are forced to conclude that $\varnothing \subseteq A$.[1]

Definition A.1.10
Set Equality

Let A, B be sets. We say that $A = B$ if $A \subseteq B$ and $B \subseteq A$.

Example A.1.11

Let

$$S = \left\{ c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ 3 \end{bmatrix} \,\middle|\, c_1, c_2, c_3 \in \mathbb{R} \right\}$$
$$T = \left\{ d_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + d_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \,\middle|\, d_1, d_2 \in \mathbb{R} \right\}.$$

Show that $S = T$.

Before we give the solution, we note that S is the set of all linear combinations of the vectors $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$ while T is the set of all linear combinations of just $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$. However, we notice that

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{A.1}$$

[1] The statement $\varnothing \subseteq A$ is called *vacuously true*, that is, it is a true statement simply because we cannot show that it is false.

Solution: We show that $S = T$ by showing that $S \subseteq T$ and that $T \subseteq S$. To show that $S \subseteq T$, we choose an arbitrary $\vec{x} \in S$ and show that $\vec{x} \in T$. So, let $\vec{x} \in S$. Then there exist $c_1, c_2, c_3 \in \mathbb{R}$ such that

$$\begin{aligned}
\vec{x} &= c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\
&= c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_3 \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \quad \text{by (A.1)} \\
&= (c_1 + c_3) \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (c_2 + c_3) \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\end{aligned}$$

from which it follows that $\vec{x} \in T$ since $\vec{x}$ can be expressed as a linear combination of $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$. This shows that $S \subseteq T$. We now show that $T \subseteq S$ by showing that if $\vec{y} \in T$ then $\vec{y} \in S$. Let $\vec{y} \in T$. Then there exist $d_1, d_2 \in \mathbb{R}$ such that

$$\begin{aligned}
\vec{y} &= d_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + d_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&= d_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + d_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ 3 \end{bmatrix}
\end{aligned}$$

from which it follows that $\vec{y} \in S$ since $\vec{y}$ can be expressed as a linear combination of $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$. Thus $T \subseteq S$. Since $S \subseteq T$ and $T \subseteq S$, we conclude that $S = T$.

Appendix B

Solutions to Exercises

This appendix contains solutions to the in-chapter exercises (but not the end-of-section problems).

1.1 Vectors in $\mathbb{R}^n$

1. No. $\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right] \neq \left[\begin{smallmatrix}2\\1\end{smallmatrix}\right]$ because their first entries are different. (And also because their second entries are different.) The order of the entries is important.

2. Rearranging gives

$$2\vec{z} = \vec{x} - 3\vec{y} = \begin{bmatrix}1\\2\\0\end{bmatrix} - 3\begin{bmatrix}-2\\1\\3\end{bmatrix} = \begin{bmatrix}7\\-1\\-9\end{bmatrix},$$

so

$$\vec{z} = \frac{1}{2}\begin{bmatrix}7\\-1\\-9\end{bmatrix} = \begin{bmatrix}7/2\\-1/2\\-9/2\end{bmatrix}$$

1.2 Linear Combinations

3.

(a) We want to find $c_1, c_2 \in \mathbb{R}$ such that

$$\begin{bmatrix}1\\-3\end{bmatrix} = c_1\begin{bmatrix}1\\-1\end{bmatrix} + c_2\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}c_1+c_2\\-c_1+c_2\end{bmatrix}.$$

By equating components, we arrive at the system of equations

$$\begin{aligned}c_1 + c_2 &= 1\\ -c_1 + c_2 &= -3.\end{aligned}$$

We can easily solve this to find that $c_1 = 2$ and $c_2 = -1$. Thus,

$$\begin{bmatrix}1\\-3\end{bmatrix} = 2\begin{bmatrix}1\\-1\end{bmatrix} + (-1)\begin{bmatrix}1\\1\end{bmatrix}.$$

(b) We proceed as in (a). We wish to find $c_1, c_2 \in \mathbb{R}$ such that

$$\begin{bmatrix} 1 \\ -3 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} c_1 + 2c_2 \\ -c_1 - 2c_2 \end{bmatrix}.$$

This time, however, the equations we obtain by equating components

$$\begin{aligned} c_1 + 2c_2 &= 1 \\ -c_1 - 2c_2 &= -3 \end{aligned}$$

do not have a solution! Indeed if we add them, we get $0 = -2$. So there are no scalars c_1 and c_2 that can be used to express $\left[\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right]$ as a linear combination of $\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 \\ -2 \end{smallmatrix}\right]$.

1.3 The Norm and the Dot Product

4. For instance, $\vec{x} = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and $\vec{y} = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ work. Another example is $\vec{z} = \left[\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right]$. Yet another one is $\vec{w} = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$. There are in fact infinitely many vectors in $\mathbb{R}^2$ of norm equal to 1. Can you describe them all?

5. If $c\vec{x}$ is a unit vector, then $\|c\vec{x}\| = 1$. On the other hand,

$$\begin{aligned} \|c\vec{x}\| &= |c|\|\vec{x}\| && \text{(by Theorem 1.3.4(b))} \\ &= |c|. && \text{(since } \vec{x} \text{ is a unit vector)} \end{aligned}$$

Putting this together, we conclude that $|c| = 1$, so $c = \pm 1$.

6. We will have

$$\vec{e}_i \cdot \vec{e}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

7. Since

$$\|\vec{x}\| = \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3},$$

From the Cauchy–Schwarz Inequality, we have that $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$, that is, that $|-7| \leq \sqrt{3}\|\vec{y}\|$. It follows that $\|\vec{y}\| \geq 7/\sqrt{3}$.

8. We have

$$\cos\theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|} = \frac{1(2) + 1(0) + 1(0) + 1(2)}{\sqrt{1^2 + 1^2 + 1^2 + 1^2}\sqrt{2^2 + 0 + 0 + 2^2}} = \frac{4}{2(2\sqrt{2})} = \frac{1}{\sqrt{2}}$$

so

$$\theta = \arccos\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{4}.$$

9. There are many such vectors (infinitely many, in fact). One of them is $\vec{x} = \left[\begin{smallmatrix} 1 \\ -1 \\ 0 \end{smallmatrix}\right]$, since $\left[\begin{smallmatrix} 1 \\ -1 \\ 0 \end{smallmatrix}\right] \cdot \left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right] = 1 - 1 + 0 = 0$. Another one is $\vec{x} = \left[\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right]$.

1.4 Vector Equations of Lines and Planes

10. A vector equation is given by

$$\vec{x} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}.$$

The above equation can be re-written as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2+t \\ -1 \\ 1+t \end{bmatrix},$$

so

$$\begin{aligned} x_1 &= 2+t \\ x_2 &= -1 \qquad t \in \mathbb{R} \\ x_3 &= 1+t \end{aligned}$$

are parametric equations for this line.

11. From the vector equation

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} + t \begin{bmatrix} 0 \\ -4 \\ -1 \end{bmatrix} = \begin{bmatrix} 1+s \\ 1+s-4t \\ -2+4s-t \end{bmatrix}, \quad s,t \in \mathbb{R}$$

we immediately get

$$\begin{aligned} x_1 &= 1+s \\ x_2 &= 1+s-4t \qquad\qquad s,t \in \mathbb{R}. \\ x_3 &= -2+4s-t \end{aligned}$$

1.5 The Cross Product in $\mathbb{R}^3$

12. We have

$$\vec{n} \cdot \vec{x} = \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1(1) + 4(2) + (-3)(3) = 0$$

and

$$\vec{n} \cdot \vec{y} = \begin{bmatrix} 1 \\ 4 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = 1(1) + 4(-1) + (-3)(-3) = 0,$$

as desired.

13. Consider $\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Then

$$(\vec{x} \times \vec{y}\,) \times \vec{w} = \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and

$$\vec{x} \times (\vec{y} \times \vec{w}) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \times \left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

so we see that $(\vec{x} \times \vec{y}) \times \vec{w} \neq \vec{x} \times (\vec{y} \times \vec{w})$. Thus, the cross product is not associative.

1.6 The Scalar Equation of Planes in $\mathbb{R}^3$

14. For $A(3,1,2)$ we simply plug $x_1 = 3$, $x_2 = 1$ and $x_3 = 2$ into the left-side of the scalar equation to obtain

$$x_1 - 3x_2 + 5x_3 = 3 - 3 + 10 = 10,$$

showing that the coordinates (x_1, x_2, x_3) satisfy the given equation. The points B and C are dealt with similarly.

15. For the direction vector we can simply take $\vec{d}_1 = \left[\begin{smallmatrix} 3 \\ -1 \\ 1 \end{smallmatrix}\right]$ (or any non-zero scalar multiple of this), and so our desired vector equation is

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + t \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}.$$

16. A normal vector for our desired plane is $\vec{n}_1 = \left[\begin{smallmatrix} 1 \\ -1 \\ 3 \end{smallmatrix}\right]$ (or any scalar multiple of this). So the desired equation is

$$(1)(x_1 - 1) + (-1)(x_2 - 0) + 3(x_3 - 0) = 0$$

which simplifies to

$$x_1 - x_2 + 3x_3 = 1.$$

1.7 Projections

17.

(a) By definition,

$$\operatorname{proj}_{\vec{v}} \vec{u} = \left(\frac{\vec{v} \cdot \vec{u}}{\|\vec{u}\|^2} \right) \vec{u}$$

and $\frac{\vec{v} \cdot \vec{u}}{\|\vec{u}\|^2}$ is a scalar.

(b) We have

$$\begin{aligned} \operatorname{proj}_{\vec{v}} \vec{u} \cdot \operatorname{perp}_{\vec{v}} \vec{u} &= \operatorname{proj}_{\vec{v}} \vec{u} \cdot (\vec{u} - \operatorname{proj}_{\vec{v}} \vec{u}) \\ &= (\operatorname{proj}_{\vec{v}} \vec{u}) \cdot u - \operatorname{proj}_{\vec{v}} \vec{u} \cdot \operatorname{proj}_{\vec{v}} \vec{u} \\ &= \left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v} \right) \cdot \vec{u} - \left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v} \right) \cdot \left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v} \right) \\ &= \left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \right) (\vec{v} \cdot \vec{u}) - \left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \right)^2 (\vec{v} \cdot \vec{v}) \end{aligned}$$

$$\begin{aligned}
&= \frac{(\vec{u}\cdot\vec{v})^2}{\|\vec{v}\|^2} - \left(\frac{(\vec{u}\cdot\vec{v})^2}{\|\vec{v}\|^4}\right)\|\vec{v}\|^2 \\
&= \frac{(\vec{u}\cdot\vec{v})^2}{\|\vec{v}\|^2} - \frac{(\vec{u}\cdot\vec{v})^2}{\|\vec{v}\|^2} \\
&= 0
\end{aligned}$$

and thus $\operatorname{proj}_{\vec{v}}\vec{u}$ and $\operatorname{perp}_{\vec{v}}\vec{u}$ are orthogonal.

(c) From the definition of the perpendicular, we see that $\operatorname{perp}_{\vec{v}}\vec{u} = \vec{u} - \operatorname{proj}_{\vec{v}}\vec{u}$, and so

$$\operatorname{proj}_{\vec{v}}\vec{u} + \operatorname{perp}_{\vec{v}}\vec{u} = \vec{u}$$

as desired.

1.8 Optional Section: Area and Volume

18. We have

$$\operatorname{area}(P) = \left\|\begin{bmatrix}1\\-1\\2\end{bmatrix}\times\begin{bmatrix}2\\-2\\4\end{bmatrix}\right\| = \left\|\begin{bmatrix}0\\0\\0\end{bmatrix}\right\| = 0.$$

The area is zero because $\vec{x}$ and $\vec{y}$ are parallel, that is, the "parallelogram" is degenerate.

19. This is just the unit cube, so its volume is 1.

2.1 Introduction and Terminology

20.

(a) Linear. Although it contains the term e^3, this is just a constant.

(b) Not linear. The appearance of the term $\frac{x_2}{x_3}$ is not permitted in a linear equation.

21. We plug $x_1 = -4$, $x_2 = 6$ and $x_3 = 1$ into the system and confirm that the equations are satisfied:

$$\begin{array}{rcrcrl}
2(-4) & + & (6) & + & 3(1) = & 1 \\
3(-4) & + & 2(6) & - & (1) = & -1 \\
5(-4) & + & 3(6) & + & 2(1) = & 0.
\end{array}$$

22. We plug $x_1 = 3-7t$, $x_2 = -5+11t$ and $x_3 = t$ into the system and confirm that the equations are satisfied:

$$\begin{array}{rcrcrl}
2(3-7t) & + & (-5+11t) & + & 3(t) = & 1 \\
3(3-7t) & + & 2(-5+11t) & - & (t) = & -1 \\
5(3-7t) & + & 3(-5+11t) & + & 2(t) = & 0.
\end{array}$$

2.2 Solving Systems of Linear Equations

23. $A = \begin{bmatrix}2&1&3\\3&2&-1\\5&3&2\end{bmatrix}$ and $\vec{b} = \begin{bmatrix}1\\-1\\0\end{bmatrix}$.

24. A is in REF but not RREF. B, C and F are in RREF (and REF). D and E are in neither.

2.3 Rank

25. REFs are given by

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

So $\operatorname{rank}(A) = 1$, $\operatorname{rank}(B) = \operatorname{rank}(C) = 2$ and $\operatorname{rank}(D) = 0$.

3.1 Matrix Algebra

26. Since

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

we require

$$\begin{array}{rcrcrcr} a & & & - & 2c & = & -3 \\ -2a & + & b & & & = & 3 \\ & & -2b & + & c & = & 6 \end{array}$$

$$\left[\begin{array}{rrr|r} 1 & 0 & -2 & -3 \\ -2 & 1 & 0 & 3 \\ 0 & -2 & 1 & 6 \end{array}\right] \xrightarrow[R_2+2R_1]{} \left[\begin{array}{rrr|r} 1 & 0 & -2 & -3 \\ 0 & 1 & -4 & -3 \\ 0 & -2 & 1 & 6 \end{array}\right] \xrightarrow[R_3+2R_2]{} \left[\begin{array}{rrr|r} 1 & 0 & -2 & -3 \\ 0 & 1 & -4 & -3 \\ 0 & 0 & -7 & 0 \end{array}\right] \xrightarrow[-\frac{1}{7}R_3]{}$$
$$\left[\begin{array}{rrr|r} 1 & 0 & -2 & -3 \\ 0 & 1 & -4 & -3 \\ 0 & 0 & 1 & 0 \end{array}\right] \xrightarrow{\substack{R_1+2R_3 \\ R_2+4R_3}} \left[\begin{array}{rrr|r} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

so $a = b = -3$ and $c = 0$.

27. We can either do this directly (using the definitions of subtraction and the transpose) or via Theorem 3.1.16. Let's use Theorem 3.1.16:

$$\begin{aligned} (A-B)^T &= (A+(-B))^T \\ &= A^T + (-B)^T && \text{by (c)} \\ &= A^T - B^T && \text{by (d)}, \end{aligned}$$

as required.

28. There are plenty. For instance, we can take $A = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$.

3.2 The Matrix–Vector Product

29.

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

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

(c) $A\vec{0} = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = 0\vec{a}_1 + \cdots + 0\vec{a}_n = \vec{0}_{\mathbb{R}^m}$.

30. By definition, we have

$$A\vec{e}_i = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_i & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = 0\vec{a}_1 + \cdots + (1)\vec{a}_i + \cdots + 0\vec{a}_n = \vec{a}_i.$$

31. There are many different possibilities here. For example, we can take $\vec{x} = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. Then $A\vec{x} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ while $B\vec{x} = \left[\begin{smallmatrix} 3 \\ 2 \end{smallmatrix}\right]$.

32.

(a) Using linear combinations, we have

$$\begin{aligned} A\vec{x} &= \begin{bmatrix} 1 & 1 & 2 & -1 \\ 2 & 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 2 \\ -3 \end{bmatrix} + 0\begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ -3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}. \end{aligned}$$

(b) Using dot products, we have

$$A\vec{x} = \begin{bmatrix} 1 & 1 & 2 & -1 \\ 2 & 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1(1) + 1(2) + 2(1) - 1(0) \\ 2(1) + 1(2) - 3(1) + 2(0) \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

3.3 The Matrix Equation $A\vec{x} = \vec{b}$

33. The system is

$$\begin{array}{rcrcr} 3x_1 & - & x_2 & = & 6 \\ 2x_1 & - & 2x_2 & = & 3 \\ -4x_1 & & & = & 2 \\ x_1 & - & 2x_2 & = & 7 \end{array}$$

34. Since $\vec{x}_1$ and $\vec{x}_2$ are solutions to $A\vec{x} = \vec{b}$, we have that $A\vec{x}_1 = A\vec{x}_2 = \vec{b}$. Then

$$\begin{aligned} A(c\vec{x}_1 + (1-c)\vec{x}_2) &= A(c\vec{x}_1) + A((1-c)\vec{x}_2) \\ &= cA\vec{x}_1 + (1-c)A\vec{x}_2 \\ &= c\vec{b} + (1-c)\vec{b} \\ &= \vec{b}. \end{aligned}$$

Thus $c\vec{x} + (1-c)\vec{x}_2$ is a solution to $A\vec{x} = \vec{b}$.

35.

(a) We have

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

which shows that $\vec{x} = \vec{s}$ is a solution of $A\vec{x} = \vec{b}$.

(b) From Theorem 3.3.9(b), we have that

$$\vec{b} = (1)\begin{bmatrix} 1 \\ 2 \end{bmatrix} + (0)\begin{bmatrix} 1 \\ 3 \end{bmatrix} + (2)\begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

3.4 Matrix Multiplication

36. We have

$$A\vec{b}_1 = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix} \quad \text{and} \quad A\vec{b}_2 = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}.$$

Therefore,

$$AB = \begin{bmatrix} 6 & 5 \\ 2 & -1 \end{bmatrix}.$$

37. We have

$$\begin{aligned} AB &= \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 2(1)+(-1)(1) & 2(1)+(-1)(2) & 2(1)+(-1)(3) & 2(1)+(-1)(4) \\ 0(1)+1(1) & 0(1)+1(2) & 0(1)+1(3) & 0(1)+1(4) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & -1 & -2 \\ 1 & 2 & 3 & 4 \end{bmatrix}. \end{aligned}$$

38. Consider *any* $A \in M_{2\times 3}(\mathbb{R})$ and $B \in M_{3\times 2}(\mathbb{R})$, for instance. Then AB is 2×2 while BA is 3×3.

39. Here are all of the defined products:

$$A_1A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A_1A_2 = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \qquad A_1A_5 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$

$$A_2A_5 = \begin{bmatrix} 14 & 11 \\ 10 & 13 \end{bmatrix} \qquad A_3A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{bmatrix} \qquad A_3A_5 = \begin{bmatrix} -4 \\ -5 \\ 3 \end{bmatrix}$$

$$A_4A_1 = \begin{bmatrix} 1 & -1 \end{bmatrix} \qquad A_4A_2 = \begin{bmatrix} -2 & 0 & 2 \end{bmatrix} \qquad A_4A_5 = \begin{bmatrix} 5 \end{bmatrix}$$

$$A_5A_4 = \begin{bmatrix} 2 & -2 \\ 3 & -3 \end{bmatrix}.$$

The remaining products are undefined for size reasons.

40. Since C commutes with both A and B, we have that $AC = CA$ and $BC = CB$. Thus

$$(AB)C = A(BC) = A(CB) = (AC)B = (CA)B = C(AB),$$

and so C commutes with AB.

41. $D^{10} = \begin{bmatrix} 2^{10} & 0 \\ 0 & (-1)^{10} \end{bmatrix} = \begin{bmatrix} 1024 & 0 \\ 0 & 1 \end{bmatrix}.$

3.5 Matrix Inverses

42. To check that a matrix B is the inverse of a matrix A, it suffices to compute the product AB and verify that it is equal to the identity matrix.

(a) We compute $(cA)(\frac{1}{c}A^{-1}) = (c\frac{1}{c})AA^{-1} = I$. Hence $(cA)^{-1} = \frac{1}{c}A^{-1}$.

(e) We compute $(A^{-1})A = I$, so $(A^{-1})^{-1} = A$.

43. We have

$$\left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 1 & 1 & -2 & 0 & 1 & 0 \\ 1 & 2 & -2 & 0 & 0 & 1 \end{array}\right] \underset{\substack{R_2-R_1 \\ R_3-R_1}}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 2 & -1 & -1 & 0 & 1 \end{array}\right] \underset{R_3-2R_2}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right] \underset{\substack{R_1+R_3 \\ R_2+R_3}}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -2 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right]$$

and we conclude that A is invertible and

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

44. The matrices $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ are not invertible because their rank is $1 < 2$, but their sum $A + B = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ is invertible. There are other examples.

45. Assume A is invertible.

(b) By the Matrix Inversion Algorithm, the RREF of A is the $n \times n$ identity matrix, which has n leading entries. Hence $\text{rank}(A) = n$.

(c) This follows from the Matrix Inversion Algorithm.

(d) If $A\vec{x} = \vec{b}$ then by multiplying both sides on the left by A^{-1} we obtain

$$A^{-1}(A\vec{x}) = A^{-1}\vec{b} \implies (A^{-1}A)\vec{x} = A^{-1}\vec{b}.$$

This shows that $\vec{x} = A^{-1}\vec{b}$ is the unique solution to the system.

(e) We claim that the inverse of A^T is given by $(A^{-1})^T$. To check this, we multiply A^T and $(A^{-1})^T$ and confirm that we get the identity matrix:

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

where in the first equality we used the fact that $(AB)^T = B^TA^T$. Thus, A^T is invertible and its inverse is $(A^{-1})^T$.

4.1 Spanning Sets

46. Simply note that $\begin{bmatrix} a \\ b \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

47. We have $\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_k$.

48. We want to determine if there are $c_1, c_2 \in \mathbb{R}$ such that

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 + 3c_2 \\ -c_1 \\ 2c_1 + c_2 \end{bmatrix}.$$

By equating entries, we obtain the system of equations

$$\begin{array}{rcrcl} c_1 & + & 3c_2 & = & 1 \\ -c_1 & & & = & 1 \\ 2c_1 & + & c_2 & = & 1 \end{array}.$$

We can solve this system by row reducing the augmented matrix, but it's quicker to see that the second equation immediately gives us $c_1 = -1$. Plugging this into the first equation gives $c_2 = 2/3$. But then the third equation is not satisfied. So we cannot find $c_1, c_2 \in \mathbb{R}$ that satisfy the system above.

We conclude that $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \notin \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\}$.

49. Let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix}$. We want to check whether $A\vec{x} = \vec{0}$ is consistent. It certainly is: $\vec{c} = \vec{0}$ is a solution. (Recall that a homogeneous system is always consistent.) Thus, $\vec{0} \in \operatorname{Span} S$, by Theorem 4.1.7.

50. Since

$$A = \begin{bmatrix} -1 & 3 & 5 \\ -1 & 1 & 1 \\ 2 & 2 & 6 \end{bmatrix} \xrightarrow[R_3+2R_1]{R_2-R_1} \begin{bmatrix} -1 & 3 & 5 \\ 0 & -2 & -4 \\ 0 & 8 & 16 \end{bmatrix} \xrightarrow[R_3+4R_2]{} \begin{bmatrix} -1 & 3 & 5 \\ 0 & -2 & -4 \\ 0 & 0 & 0 \end{bmatrix},$$

we see that $\operatorname{rank}(A) = 2 < 3$, so S does not span $\mathbb{R}^3$ by Theorem 4.1.10.

51. Let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix} \in M_{n\times k}(\mathbb{R})$. Then

$$\operatorname{rank}(A) \le \min\{k, n\} = k < n.$$

Thus S cannot span $\mathbb{R}^n$ by Theorem 4.1.10.

52. Consider the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

Letting

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

we see that $\operatorname{rank}(A) = 1 < 3$, so S does not span $\mathbb{R}^3$ by Theorem 4.1.10.

4.2 Geometry of Spanning Sets

53. By definition,
$$U = \operatorname{Span}\{\vec{v}_1\} = \{c_1\vec{v}_1 \mid c_1 \in \mathbb{R}\}.$$
Thus, $\vec{x} \in U$ if and only if it satisfies $\vec{x} = c_1\vec{v}_1$ for some $c_1 \in \mathbb{R}$. Since $\vec{v}_1 \neq \vec{0}$, we recognize $\vec{x} = c_1\vec{v}_1$ as the vector equation of a line. Hence, U is a line in $\mathbb{R}^3$ through the origin.

54. If $\vec{v}_2 = c\vec{v}_1$, then every linear combination of $\vec{v}_1$ and $\vec{v}_2$ will be a scalar multiple of $\vec{v}_1$:
$$a\vec{v}_1 + b\vec{v}_2 = a\vec{v}_1 + (bc)\vec{v}_1 = (a+bc)\vec{v}_1.$$
From this we deduce that $U = \operatorname{Span}\{\vec{v}_1, \vec{v}_2\} = \operatorname{Span}\{\vec{v}_1\}$.

Thus, if $\vec{v}_1 = \vec{0}$, we would have $U = \operatorname{Span}\{\vec{0}\} = \{\vec{0}\}$. Otherwise, if $\vec{v}_1 \neq \vec{0}$, U is a line through the origin with direction vector $\vec{v}_1$.

55. We solve the system $A\vec{x} = \vec{v}$ where
$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$
We have
$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & v_1 \\ 0 & 1 & 2 & v_2 \\ 0 & 0 & 1 & v_3 \end{array}\right] \xrightarrow[\substack{R_1 - R_3 \\ R_2 - 2R_3}]{} \left[\begin{array}{ccc|c} 1 & 1 & 0 & v_1 - v_3 \\ 0 & 1 & 0 & v_2 - 2v_3 \\ 0 & 0 & 1 & v_3 \end{array}\right] \xrightarrow{R_1 - R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & v_1 - v_2 + v_3 \\ 0 & 1 & 0 & v_2 - 2v_3 \\ 0 & 0 & 1 & v_3 \end{array}\right].$$
Thus
$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = (v_1 - v_2 + v_3)\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (v_2 - 2v_3)\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + v_3\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

56. We have
$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix},$$
so we can discard $\left[\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right]$ from the spanning set, leaving us with
$$U = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -5 \\ -5 \\ 3 \end{bmatrix}\right\}.$$
Next, we have
$$\begin{bmatrix} -5 \\ -5 \\ 3 \end{bmatrix} = (-5)\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{3}{2}\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$
so we can discard $\left[\begin{smallmatrix} -5 \\ -5 \\ 3 \end{smallmatrix}\right]$, leaving us with
$$U = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}\right\}.$$
We cannot simplify U further since neither $\left[\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right]$ nor $\left[\begin{smallmatrix} 0 \\ 0 \\ 2 \end{smallmatrix}\right]$ is a scalar multiples of the other.

57.

(a) Recall that

$$U = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

We have already seen in Example 4.2.7 that

$$S_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a spanning set for U and that neither vector in S_1 is a scalar multiple of the other, showing that U is a plane through the origin in $\mathbb{R}^3$. Since

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

The Reduction Theorem shows that

$$S_2 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a spanning set for U. Since neither of the vectors in S_2 is a linear combination of the other, we cannot further reduce this spanning set. Finally, since

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

The Reduction Theorem again shows that

$$S_3 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a spanning set for U. Since neither of the vectors in S_3 is a linear combination of the other, we again cannot further reduce this spanning set.

(b) Since all subsets of S containing 2 vectors span U and none of these subsets can be further reduced, there are no subsets of S containing just 1 vector that are spanning sets for U. Alternatively, since a set containing one vector can span at most a line and U is a plane, no subset of S containing just one vector can span U.

58. Using the Reduction Theorem, we have

$$\begin{aligned} &\text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \right\} \\ &= \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} \right\} \qquad \text{since } \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

$$= \text{Span}\left\{\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}2\\1\\1\end{bmatrix}, \begin{bmatrix}3\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}\right\} \qquad \text{since } \begin{bmatrix}-1\\0\\1\end{bmatrix} = 1\begin{bmatrix}1\\1\\2\end{bmatrix} - 1\begin{bmatrix}2\\1\\1\end{bmatrix} + 0\begin{bmatrix}3\\2\\3\end{bmatrix}$$

$$= \text{Span}\left\{\begin{bmatrix}2\\1\\1\end{bmatrix}, \begin{bmatrix}3\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}\right\} \qquad \text{since } \begin{bmatrix}1\\1\\2\end{bmatrix} = -1\begin{bmatrix}2\\1\\1\end{bmatrix} + \begin{bmatrix}3\\2\\3\end{bmatrix} + 0\begin{bmatrix}-1\\0\\1\end{bmatrix}$$

$$= \text{Span}\left\{\begin{bmatrix}3\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}\right\} \qquad \text{since } \begin{bmatrix}2\\1\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}3\\2\\3\end{bmatrix} - \frac{1}{2}\begin{bmatrix}-1\\0\\1\end{bmatrix}.$$

4.3 Linear Dependence and Linear Independence

59. For $c_1, c_2 \in \mathbb{R}$, consider

$$c_1\begin{bmatrix}1\\0\\1\end{bmatrix} + c_2\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

We obtain the system of equations

$$\begin{array}{rcrcl} c_1 & + & c_2 & = & 0 \\ & & c_2 & = & 0 \\ -c_1 & + & c_2 & = & 0 \end{array}.$$

Carrying the coefficient matrix to row echelon form gives

$$\begin{bmatrix}1 & 1\\0 & 1\\-1 & 1\end{bmatrix} \underset{R_3+R_1}{\longrightarrow} \begin{bmatrix}1 & 1\\0 & 1\\0 & 2\end{bmatrix} \underset{R_3-2R_2}{\longrightarrow} \begin{bmatrix}1 & 1\\0 & 1\\0 & 0\end{bmatrix},$$

from which we see that there are no free variables and hence a unique (trivial) solution. Thus S is linearly independent.

In fact, we see from the second equation that $c_2 = 0$ and substituting $c_2 = 0$ into both the first and third equations each gives $c_1 = 0$. Thus we have only the trivial solution $c_1 = c_2 = 0$ and we conclude, again, that S is linearly independent.

60. The set S is linearly dependent. Consider

$$c_1\begin{bmatrix}1\\2\end{bmatrix} + c_2\begin{bmatrix}1\\3\end{bmatrix} + c_3\begin{bmatrix}1\\4\end{bmatrix} + c_4\begin{bmatrix}1\\5\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.$$

We obtain the system of equations

$$\begin{array}{rcrcrcrcl} c_1 & + & c_2 & + & c_3 & + & c_4 & = & 0 \\ 2c_1 & + & 3c_2 & + & 4c_3 & + & 5c_4 & = & 0 \end{array}.$$

Carrying the coefficient matrix to row echelon form gives

$$\begin{bmatrix}1 & 1 & 1 & 1\\2 & 3 & 4 & 5\end{bmatrix} \underset{R_2-2R_1}{\longrightarrow} \begin{bmatrix}1 & 1 & 1 & 1\\0 & 1 & 2 & 3\end{bmatrix}.$$

From this we see that the rank of the coefficient matrix is 2, and that there are two free variables. In particular, the system has non-trivial solutions. Thus, the set S must be linearly dependent, as claimed.

With a little more work, we can find all non-trivial solutions. One of them is $c_1 = 1$, $c_2 = -2$, $c_3 = 1$ and $c_4 = 0$.

61. By Theorem 4.3.4, the set S will be linearly independent if and only if $\operatorname{rank}(A) = 3$. We have

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \underset{\substack{R_2-2R_1 \\ R_3-3R_1}}{\longrightarrow} \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix} \underset{R_3-2R_2}{\longrightarrow} \begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}.$$

We thus see that $\operatorname{rank}(A) = 2 < 3$, so S is linearly dependent.

62. Let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_k \end{bmatrix} \in M_{n\times k}(\mathbb{R})$. Then

$$\operatorname{rank}(A) \leq \min\{k, n\} = n < k.$$

Thus S is linearly dependent by Theorem 4.3.4.

63. Consider the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

Letting

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

we see that $\operatorname{rank}(A) = 1 < 3$, so S is linearly dependent by Theorem 4.3.4.

64. For $c_1, c_2, c_3 \in \mathbb{R}$, consider

$$c_1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix} + c_3 \begin{bmatrix} -1 \\ 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{B.1}$$

Carrying the coefficient matrix of this homogeneous system of linear equations to reduced row echelon form gives

$$\begin{bmatrix} 1 & -3 & -1 \\ 2 & 2 & 6 \\ 1 & 4 & 6 \end{bmatrix} \underset{\substack{R_2-2R_1 \\ r_3-R_1}}{\longrightarrow} \begin{bmatrix} 1 & -3 & -1 \\ 0 & 8 & 8 \\ 0 & 7 & 7 \end{bmatrix} \underset{\frac{1}{8}R_2}{\longrightarrow} \begin{bmatrix} 1 & -3 & -1 \\ 0 & 1 & 1 \\ 0 & 7 & 7 \end{bmatrix} \underset{R_3-7R_2}{\overset{R_1+3R_2}{\longrightarrow}} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

As the rank of the coefficient matrix is $2 < 3$, S is linearly dependent and so the Dependency Theorem guarantees that at least one vector in S is a linear combination of the others. Solving the above system gives

$$c_1 = -2t, \quad c_2 = -t \quad \text{and} \quad c_3 = t$$

for any $t \in \mathbb{R}$. Taking $t = 1$, it follows from (B.1) that

$$-2 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 1 \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

We see that each vector from S can be expressed as a linear combination of the other vectors. Using the Reduction Theorem, we have

$$\begin{bmatrix}1\\2\\1\end{bmatrix} = -\frac{1}{2}\begin{bmatrix}-3\\2\\4\end{bmatrix} + \frac{1}{2}\begin{bmatrix}-1\\6\\6\end{bmatrix} \implies \operatorname{Span}\left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-3\\2\\4\end{bmatrix}, \begin{bmatrix}-1\\6\\6\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}-3\\2\\4\end{bmatrix}, \begin{bmatrix}-1\\6\\6\end{bmatrix}\right\}$$

$$\begin{bmatrix}-3\\2\\4\end{bmatrix} = -2\begin{bmatrix}1\\2\\1\end{bmatrix} + 1\begin{bmatrix}-1\\6\\6\end{bmatrix} \implies \operatorname{Span}\left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-3\\2\\4\end{bmatrix}, \begin{bmatrix}-1\\6\\6\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-1\\6\\6\end{bmatrix}\right\}$$

$$\begin{bmatrix}-1\\6\\6\end{bmatrix} = 2\begin{bmatrix}1\\2\\1\end{bmatrix} + 1\begin{bmatrix}-3\\2\\6\end{bmatrix} \implies \operatorname{Span}\left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-3\\2\\4\end{bmatrix}, \begin{bmatrix}-1\\6\\6\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-3\\2\\4\end{bmatrix}\right\}.$$

Thus each of the subsets

$$\left\{\begin{bmatrix}-3\\2\\4\end{bmatrix}, \begin{bmatrix}-1\\6\\6\end{bmatrix}\right\}, \quad \left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-1\\6\\6\end{bmatrix}\right\} \quad \text{and} \quad \left\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}-3\\2\\4\end{bmatrix}\right\}$$

are spanning sets for U. Since in any given set, neither vector is a scalar multiple of the other, we have obtained all linearly independent subsets of S that are spanning sets for U. We thus recognize U as a plane through the origin in $\mathbb{R}^3$.

65. We carry

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

to row echelon form. We have

$$\begin{bmatrix}0&1&2&0&1&0\\0&1&2&1&4&0\\0&-1&-2&0&-1&1\\0&2&4&2&8&0\end{bmatrix} \underset{\substack{R_2-R_2\\R_3+R_1\\R_4-2R_1}}{\longrightarrow} \begin{bmatrix}0&1&2&0&1&0\\0&0&0&1&3&0\\0&0&0&0&0&1\\0&0&0&2&6&0\end{bmatrix} \underset{R_4-2R_2}{\longrightarrow} \begin{bmatrix}0&1&2&0&1&0\\0&0&0&1&3&0\\0&0&0&0&0&1\\0&0&0&0&0&0\end{bmatrix}.$$

Since a row echelon form of A has leading entries in the second, fourth and sixth columns, the Extraction Theorem gives that

$$S' = \left\{\begin{bmatrix}1\\1\\-1\\2\end{bmatrix}, \begin{bmatrix}0\\1\\0\\2\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}\right\}$$

is a linearly independent subset of S with $U = \operatorname{Span} S'$.

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

66. Properties S1, S2 and S3 are all trivially satisfied. Indeed, U if nonempty and $\vec{0} + \vec{0} = \vec{0}$ and $c\vec{0} = \vec{0}$ for all $c \in \mathbb{R}$.

67. We note that $\vec{0} \notin U$ because

$$0 - 0 + 2(0) \neq 4.$$

You can also check that U is not closed under addition or scalar multiplication.

4.5 Bases and Dimension

68.

(a) Since

$$A_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \underset{R_3 - R_1}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \underset{R_3 - R_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \end{bmatrix},$$

we see that $\operatorname{rank}(A_1) = 3$, so B_1 is a basis for $\mathbb{R}^3$ by Theorem 4.5.7.

(b) Since

$$A_2 = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 1 & 3 \\ 3 & 2 & 7 \end{bmatrix} \underset{\substack{R_2 - R_1 \\ R_3 - 3R_1}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 5 \\ 0 & -1 & -2 \\ 0 & -4 & -8 \end{bmatrix} \underset{R_3 - 4R_2}{\longrightarrow} \begin{bmatrix} 1 & 2 & 5 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{bmatrix},$$

we see that $\operatorname{rank}(A_2) = 2 < 3$, so B_2 is not a basis for $\mathbb{R}^3$ by Theorem 4.5.7.

69. Let $A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} \in M_{n \times n}(\mathbb{R})$. Then

$$\begin{aligned} B \text{ spans } \mathbb{R}^n &\iff \operatorname{rank}(A) = n && \text{by Theorem 4.1.10} \\ &\iff B \text{ is linearly independent} && \text{by Theorem 4.3.4.} \end{aligned}$$

70. Let $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right] \in U$. Then $x_1 + 2x_2 = 0$, so $x_1 = -2x_2$. We have that

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

Thus $U = \operatorname{Span}\left\{ \left[\begin{smallmatrix} -2 \\ 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right] \right\}$. Thus the set

$$B = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a spanning set for U. Since neither vector in B is a scalar multiple of the other, B is linearly independent, and hence a basis for U.

71.

(a) We have

$$\begin{bmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 3 & 2 \\ 0 & 1 & -1 & -1 \end{bmatrix} \underset{R_3 - R_1}{\longrightarrow} \begin{bmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix} \underset{\substack{R_3 + R_2 \\ R_4 - R_2}}{\longrightarrow} \begin{bmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -2 \end{bmatrix} \underset{R_4 + 2R_3}{\longrightarrow} \begin{bmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since the last matrix is in row echelon form with leading entries in the first, second and fourth columns, the Extraction Theorem gives that

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \\ -1 \end{bmatrix} \right\}$$

is a basis for U.

(b) Let A be the matrix whose first three columns are the vectors in B and whose last four columns are the standard basis vectors $\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4$:

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

We carry A to row echelon form:

$$\begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \underset{R_3 - R_1}{\longrightarrow} \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \underset{\substack{R_3 + R_2 \\ R_4 - R_2}}{\longrightarrow}$$

$$\begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 & 0 \\ 0 & 0 & -2 & 0 & -1 & 0 & 1 \end{bmatrix} \underset{R_4 + 2R_3}{\longrightarrow} \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 & 0 \\ 0 & 0 & 0 & -2 & 1 & 2 & 1 \end{bmatrix}.$$

Since the above row echelon of A has leading entries in the first four columns, it follows from the Extraction Theorem that the first four columns of A will form our desired basis for $\mathbb{R}^4$. Thus we have

$$B' = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

4.6 Fundamental Subspaces Associated with a Matrix

72. We have

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 \\ 3 & 5 & 2 & 4 \end{bmatrix} \underset{\substack{R_2 - 2R_1 \\ R_3 - 3R_1}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & -1 & -1 & -2 \\ 0 & -1 & -1 & -2 \end{bmatrix} \underset{R_3 - R_2}{\overset{R_1 + 2R_2}{\longrightarrow}} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & -1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \underset{-R_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

so the solution to $A\vec{x} = \vec{0}$ is

$$\vec{x} = s \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R}.$$

and thus

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathrm{Null}(A)$.

73. We have

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 \\ 3 & 5 & 2 & 4 \end{bmatrix} \underset{\substack{R_2 - 2R_1 \\ R_3 - 3R_1}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & -1 & -1 & -2 \\ 0 & -1 & -1 & -2 \end{bmatrix} \underset{R_3 - R_2}{\overset{R_1 + 2R_2}{\longrightarrow}} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & -1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since there are leading entries in the first and second columns of a row echelon form of A, the first and second columns of A will form a basis for $\mathrm{Col}(A)$. Thus

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \right\}$$

is a basis for $\mathrm{Col}(A)$.

5.1 Matrix Transformations and Linear Transformations

74. The domain is $\mathbb{R}^2$, the codomain is $\mathbb{R}^3$, and we have

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

75. Consider $\vec{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and $c = 2$. Then

$$T(c\vec{x}) = T\left(\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 0+0+2 \\ 2^2+3 \end{bmatrix} = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

while

$$cT(\vec{x}) = 2T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = 2\begin{bmatrix} 1 \\ 1^2+3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}.$$

Thus, $T(c\vec{x}) \neq cT(\vec{x})$.

76. Consider

$$\vec{x} = \vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{y} = \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Then

$$T(\vec{x}+\vec{y}) = T(\vec{e}_1+\vec{e}_2) = T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \left\|\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\| = \sqrt{2},$$

and

$$T(\vec{x}) + T(\vec{y}) = T(\vec{e}_1) + T(\vec{e}_2) = \|\vec{e}_1\| + \|ve_2\| = 1+1 = 2.$$

Hence $T(\vec{x}+\vec{y}) \neq T(\vec{x}) + T(\vec{y})$.

77. We have

$$\left[T\right] = \left[T(\vec{e}_1) \;\; T(\vec{e}_2)\right] = \left[T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \;\; T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)\right] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 3 \end{bmatrix}.$$

78. Our first goal is to express the standard basis vectors $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ as linear combinations of $\left[\begin{smallmatrix} 1 \\ 0 \\ 1 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 \\ 1 \\ 1 \end{smallmatrix}\right]$.

We can do this by setting up and solving the relevant systems of linear equations (namely, the systems with augmented matrices $[A \mid \vec{e}_i]$ for $i = 1, 2, 3$), but in this case it's easier to do this by inspection! We have

$$\begin{aligned}
\vec{e}_1 &= \begin{bmatrix}1\\0\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\0\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}\\
\vec{e}_2 &= \begin{bmatrix}0\\1\\0\end{bmatrix} = -\frac{1}{2}\begin{bmatrix}1\\0\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix} + \begin{bmatrix}0\\1\\1\end{bmatrix}\\
\vec{e}_3 &= \begin{bmatrix}0\\0\\1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}.
\end{aligned}$$

It follows that

$$\begin{aligned}
T(\vec{e}_1) &= \frac{1}{2}T\left(\begin{bmatrix}1\\0\\1\end{bmatrix}\right) + \frac{1}{2}T\left(\begin{bmatrix}1\\0\\-1\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}2\\4\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\2\\1\end{bmatrix} = \begin{bmatrix}1\\3\\1\end{bmatrix}\\
T(\vec{e}_2) &= -\frac{1}{2}T\left(\begin{bmatrix}1\\0\\1\end{bmatrix}\right) + \frac{1}{2}T\left(\begin{bmatrix}1\\0\\-1\end{bmatrix}\right) + T\left(\begin{bmatrix}0\\1\\1\end{bmatrix}\right) = -\frac{1}{2}\begin{bmatrix}2\\4\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\2\\1\end{bmatrix} + \begin{bmatrix}-1\\1\\3\end{bmatrix}\\
&= \begin{bmatrix}-2\\0\\3\end{bmatrix}\\
T(\vec{e}_3) &= \frac{1}{2}T\left(\begin{bmatrix}1\\0\\1\end{bmatrix}\right) - \frac{1}{2}T\left(\begin{bmatrix}1\\0\\-1\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}2\\4\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}0\\2\\1\end{bmatrix} = \begin{bmatrix}1\\1\\0\end{bmatrix}.
\end{aligned}$$

Thus,

$$\big[T\big] = \big[T(\vec{e}_1) \;\; T(\vec{e}_2) \;\; T(\vec{e}_3)\big] = \begin{bmatrix}1 & -2 & 1\\3 & 0 & 1\\1 & 3 & 0\end{bmatrix}.$$

5.2 Examples of Linear Transformations

79. Let $\vec{x}, \vec{y} \in \mathbb{R}^2$ and $c_1, c_2 \in \mathbb{R}$. Then

$$\begin{aligned}
S(c_1\vec{x} + c_2\vec{y}) &= \operatorname{perp}_{\vec{d}}(c_1\vec{x} + c_2\vec{y})\\
&= (c_1\vec{x} + c_2\vec{y}) - \operatorname{proj}_{\vec{d}}(c_1\vec{x} + c_2\vec{y})\\
&= (c_1\vec{x} + c_2\vec{y}) - (c_1 \operatorname{proj}_{\vec{d}} \vec{x} + c_2 \operatorname{proj}_{\vec{d}} \vec{y}) && \text{by Theorem 5.2.1(a)}\\
&= c_1(\vec{x} - \operatorname{proj}_{\vec{d}} \vec{x}) + c_2(\vec{y} - \operatorname{proj}_{\vec{d}} \vec{y})\\
&= c_1 \operatorname{perp}_{\vec{d}} \vec{x} + c_2 \operatorname{perp}_{\vec{d}} \vec{y}\\
&= c_1 S(\vec{x}) + c_2 S(\vec{y}),
\end{aligned}$$

so S is linear by the Linearity Test.

80. With $\vec{n} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$,

$$T(\vec{e}_1) = \operatorname{perp}_{\vec{n}} \vec{e}_1 = \vec{e}_1 - \operatorname{proj}_{\vec{n}} \vec{e}_1 = \vec{e}_1 - \frac{\vec{e}_1 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{2}{6}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ -1/3 \end{bmatrix}$$

$$T(\vec{e}_2) = \operatorname{perp}_{\vec{n}} \vec{e}_2 = \vec{e}_2 - \operatorname{proj}_{\vec{n}} \vec{e}_2 = \vec{e}_2 - \frac{\vec{e}_2 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{-1}{6}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 5/6 \\ 1/6 \end{bmatrix}$$

$$T(\vec{e}_3) = \operatorname{perp}_{\vec{n}} \vec{e}_3 = \vec{e}_3 - \operatorname{proj}_{\vec{n}} \vec{e}_3 = \vec{e}_3 - \frac{\vec{e}_3 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{6}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/3 \\ 1/6 \\ 5/6 \end{bmatrix}$$

so

$$\left[T\right] = \left[T(\vec{e}_1) \;\; T(\vec{e}_2) \;\; T(\vec{e}_3)\right] = \begin{bmatrix} 1/3 & 1/3 & -1/3 \\ 1/3 & 5/6 & 1/6 \\ -1/3 & 1/6 & 5/6 \end{bmatrix}.$$

81. The standard matrix is $\begin{bmatrix} 1 & 0 \\ 0 & t \end{bmatrix}$.

82. The standard matrix is $\begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}$ with $s = 3$. So

$$T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}.$$

5.3 Operations on Linear Transformations

83.

(a) For $\vec{x}, \vec{y} \in \mathbb{R}^n$, we have

$$T(\vec{x} + \vec{y}) = \vec{0} = \vec{0} + \vec{0} = T(\vec{x}) + T(\vec{y}).$$

and for $c \in \mathbb{R}$,

$$T(c\vec{x}) = \vec{0} = c\vec{0} = cT(\vec{x}).$$

(b) We have that

$$\left[T\right] = \left[T(\vec{e}_1) \;\; \cdots \;\; T(\vec{e}_n)\right] = \left[\vec{0} \;\; \cdots \;\; \vec{0}\right] = 0_{m\times n}.$$

84.

(a) Theorem 5.3.7 tells us that cT and dS are linear, since T and S are linear. And so their sum $(cT) + (dS)$ must be linear too (again by Theorem Theorem 5.3.7).

(b) By appealing to Theorem 5.3.7 one more time, we find that

$$\left[cT + dS\right] = \left[cT\right] + \left[dS\right] = c\left[T\right] + d\left[S\right],$$

as required.

85.

(a) $$S\left(T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)\right) = S\left(\begin{bmatrix}x_2\\x_1+x_2\\x_1\end{bmatrix}\right) = \begin{bmatrix}x_2+(x_1+x_2)+x_1\\x_2+(x_1+x_2)-x_1\end{bmatrix} = \begin{bmatrix}2(x_1+x_2)\\2x_2\end{bmatrix}.$$

(b) $$T\left(S\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right)\right) = T\left(\begin{bmatrix}x_1+x_2+x_3\\x_1+x_2-x_3\end{bmatrix}\right) = \begin{bmatrix}x_1+x_2-x_3\\(x_1+x_2-x_3)+(x_1+x_2+x_3)\\x_1+x_2+x_3\end{bmatrix}.$$

86. We have

$$\big[T\big] = \big[\operatorname{proj}_{\vec{d}}(\vec{e}_1)\ \ \operatorname{proj}_{\vec{d}}(\vec{e}_2)\big] = \begin{bmatrix}1/2 & 1/2\\1/2 & 1/2\end{bmatrix}$$

hence

$$\big[T\circ T\big] = \big[T\big]\,\big[T\big] = \begin{bmatrix}1/2 & 1/2\\1/2 & 1/2\end{bmatrix}\begin{bmatrix}1/2 & 1/2\\1/2 & 1/2\end{bmatrix} = \begin{bmatrix}1/2 & 1/2\\1/2 & 1/2\end{bmatrix}.$$

Since $\big[T\circ T\big] = \big[T\big]$, it follows from Theorem 5.3.4 that $T\circ T = T$.

5.4 Inverses of Linear Transformations

87. We must simply show that $[T]^{-1} = [S]$. We have

$$\big[T\big] = \begin{bmatrix}2 & 3\\1 & 2\end{bmatrix} \quad\text{and}\quad \big[S\big] = \begin{bmatrix}2 & -3\\-1 & 2\end{bmatrix}.$$

We compute

$$\big[T\big]\,\big[S\big] = \begin{bmatrix}2 & 3\\1 & 2\end{bmatrix}\begin{bmatrix}2 & -3\\-1 & 2\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}.$$

This shows that $\big[S\big] = \big[T\big]^{-1}$, which is what we wanted to show.

88. Since reflecting through a plane in $\mathbb{R}^3$ can be "undone" by performing the reflection again, we have that $T^{-1} = T$. Thus

$$\big[T\big]^{-1} = \big[T^{-1}\big] = \big[T\big] = \begin{bmatrix}2/3 & 1/3 & -2/3\\1/3 & 2/3 & 2/3\\-2/3 & 2/3 & -1/3\end{bmatrix}.$$

89. Since

$$\big[T\big] = \big[T(\vec{e}_1)\ \ T(\vec{e}_2)\ \ T(\vec{e}_3)\big] = \begin{bmatrix}1 & 0 & -1\\1 & 1 & 0\\1 & 1 & 1\end{bmatrix},$$

the Matrix Inversion Algorithm gives

$$\left[\begin{array}{ccc|ccc}1 & 0 & -1 & 1 & 0 & 0\\1 & 1 & 0 & 0 & 1 & 0\\1 & 1 & 1 & 0 & 0 & 1\end{array}\right] \xrightarrow[R_3-R_1]{R_2-R_1} \left[\begin{array}{ccc|ccc}1 & 0 & -1 & 1 & 0 & 0\\0 & 1 & 1 & -1 & 1 & 0\\0 & 1 & 2 & -1 & 0 & 1\end{array}\right] \xrightarrow[R_3-R_2]{}$$

$$\left[\begin{array}{ccc|ccc}1 & 0 & -1 & 1 & 0 & 0\\0 & 1 & 1 & -1 & 1 & 0\\0 & 0 & 1 & 0 & -1 & 1\end{array}\right] \xrightarrow{\substack{R_1+R_3\\R_2-R_3}} \left[\begin{array}{ccc|ccc}1 & 0 & 0 & 1 & -1 & 1\\0 & 1 & 0 & -1 & 2 & -1\\0 & 0 & 1 & 0 & -1 & 1\end{array}\right].$$

Thus,

$$\left[T^{-1}\right] = \left[T\right]^{-1} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

so

$$T^{-1}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 + x_3 \\ -x_1 + 2x_2 - x_3 \\ -x_2 + x_3 \end{bmatrix}.$$

5.5 The Kernel and the Range

90. Since

$$T_1\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 2(1) - 2 \\ 3(2) - 2(3) \\ 1 + 2 - 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right] \in \mathrm{Ker}(T_1)$. We next compute

$$T_2\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 1 - 5(2) + 4(3) \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which shows that $\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right] \notin \mathrm{Ker}(T_2)$. Finally,

$$T_3\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = 5(1) - 4(2) + 3 = 0,$$

so $\left[\begin{smallmatrix} 1 \\ 2 \\ 3 \end{smallmatrix}\right] \in \mathrm{Ker}(T_3)$.

91. Consider first $T_1\left(\left[\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 1 \\ 3 \\ 3 \end{smallmatrix}\right]$. This leads to the system of linear equations

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

whose augmented matrix we carry to row echelon form. We have

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 \\ -2 & 0 & -1 & 3 \end{array}\right] \underset{R_3+2R_1}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 \\ 0 & 2 & -3 & 5 \end{array}\right] \underset{R_3-2R_2}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

which shows the system is consistent and so $\left[\begin{smallmatrix} 1 \\ 3 \\ 3 \end{smallmatrix}\right] \in \mathrm{Range}(T_1)$. Note that by solving the above system, we will find that $\vec{x} = \left[\begin{smallmatrix} -1 \\ 1 \\ -1 \end{smallmatrix}\right]$, that is, we will find that $T_1\left(\left[\begin{smallmatrix} -1 \\ 1 \\ -1 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 1 \\ 3 \\ 3 \end{smallmatrix}\right]$.

We now consider $T_2\left(\left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 1 \\ 3 \\ 3 \end{smallmatrix}\right]$. Carrying the augmented matrix of the resulting system to row echelon form gives

$$\left[\begin{array}{cc|c} 2 & -3 & 1 \\ 1 & 1 & 3 \\ 2 & -1 & 3 \end{array}\right] \overset{R_1\leftrightarrow R_2}{\longrightarrow} \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 2 & -3 & 1 \\ 2 & -1 & 3 \end{array}\right] \underset{\substack{R_2-2R_1 \\ R_3-2R_1}}{\longrightarrow} \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & -5 & -5 \\ 0 & -3 & -3 \end{array}\right] \underset{R_3-\frac{3}{5}R_2}{\longrightarrow} \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & -5 & -5 \\ 0 & 0 & 0 \end{array}\right]$$

from which we see the system is consistent, showing that $\begin{bmatrix} 1\\3\\3 \end{bmatrix} \in \text{Range}(T_2)$. Note that by solving the above system, we will find that $\vec{x} = \left[\begin{smallmatrix} 2\\1 \end{smallmatrix}\right]$, so $T_2\left(\left[\begin{smallmatrix} 2\\1 \end{smallmatrix}\right]\right) = \begin{bmatrix} 1\\3\\3 \end{bmatrix}$.

92.

(a) Since T is linear, $T(\vec{0}_{\mathbb{R}^n}) = \vec{0}_{\mathbb{R}^m}$ so $\vec{0}_{\mathbb{R}^n} \in \text{Ker}(T)$. For $\vec{x}, \vec{y} \in \text{Ker}(T)$, we have that $T(\vec{x}) = \vec{0} = T(\vec{y})$. Then, since T is linear

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}) = \vec{0} + \vec{0} = \vec{0}$$

so $\vec{x} + \vec{y} \in \text{Ker}(T)$ and $\text{Ker}(T)$ is closed under vector addition. For $c \in \mathbb{R}$, we again use the linearity of T to obtain

$$T(c\vec{x}) = cT(\vec{x}) = c\,\vec{0} = \vec{0}$$

showing that $c\vec{x} \in \text{Ker}(T)$ so that $\text{Ker}(T)$ is closed under scalar multiplication. Hence, $\text{Ker}(T)$ is a subspace of $\mathbb{R}^n$.

(b) Since T is linear, $T(\vec{0}_{\mathbb{R}^n}) = \vec{0}_{\mathbb{R}^m}$ so $\vec{0}_{\mathbb{R}^m} \in \text{Range}(T)$. For $\vec{x}, \vec{y} \in \text{Range}(T)$, there exist $\vec{u}, \vec{v} \in \mathbb{R}^n$ such that $\vec{x} = T(\vec{u})$ and $\vec{y} = T(\vec{\ })$. Then since T is linear,

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}) = \vec{x} + \vec{y}$$

and so $\vec{x} + \vec{y} \in \text{Range}(T)$. For $c \in \mathbb{R}$, we use the linearity of T to obtain

$$T(c\vec{u}) = cT(\vec{u}) = c\vec{x}$$

and so $c\vec{x} \in \text{Range}(T)$. Thus $\text{Range}(T)$ is a subspace of $\mathbb{R}^m$.

93. We have

$$\big[\,T\,\big] = \big[\,T(\vec{e}_1) \;\; T(\vec{e}_2) \;\; T(\vec{e}_3)\,\big] = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Carrying $\big[\,T\,\big]$ to reduced row echelon form gives

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \underset{R_2 - R_1}{\longrightarrow} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

from which we see the solution to $T(\vec{x}) = \big[\,T\,\big]\,\vec{x} = \vec{0}$ is

$$\vec{x} = t \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}$$

and so

$$\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis for $\text{Ker}(T)$. As the reduced row echelon form of $\big[\,T\,\big]$ has leading ones in the first and last columns, a basis for $\text{Range}(T)$ is

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

94. If $\text{Range}(T) = \mathbb{R}^m$, then $\dim(\text{Range}(T)) = \dim(\mathbb{R}^m) = m$. Thus Theorem 5.5.7 gives that

$$n = \dim(\text{Ker}(T)) + \dim(\text{Range}(T)) = \dim(\text{Ker}(T)) + m$$

from which it follows that $m = n - \dim(\text{Ker}(T))$. Since $\dim(\text{Ker}(T)) \geq 0$, we see that

$$m = n - \dim(\text{Ker}(T)) \leq n - 0 = n.$$

Thus, $m \leq n$.

6.1 Determinants and Invertibility

95. We have

$$\det(A) = \det\left(\begin{bmatrix} 2 & -1 \\ 5 & -3 \end{bmatrix}\right) = 2(-3) - (-1)5 = -6 + 5 = -1.$$

Since $\det(A) \neq 0$, A is invertible. Next,

$$\det(B) = \det\left(\begin{bmatrix} 3 & -6 \\ -1 & 2 \end{bmatrix}\right) = 3(2) - (-6)(-1) = 6 - 6 = 0.$$

Since $\det(B) = 0$, B is not invertible.

96. We have

$$\begin{aligned}
C_{11}(A) &= (-1)^{1+1}\det(A(1,1)) = (-1)^2 \det\left(\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\right) = 1(1-4) = -3, \\
C_{21}(A) &= (-1)^{2+1}\det(A(2,1)) = (-1)^3 \det\left(\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}\right) = -1(1-(-2)) = -3, \\
C_{31}(A) &= (-1)^{3+1}\det(A(3,1)) = (-1)^4 \det\left(\begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}\right) = 1(2-(-1)) = 3.
\end{aligned}$$

97. Performing a cofactor expansion along the first row of A gives

$$\begin{aligned}
\det(A) = \det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right)& \\
&= aC_{11}(A) + bC_{12}(A) + cC_{13}(A) \\
&= a(-1)^{1+1}\det(A(1,1)) + b(-1)^{1+2}\det(A(1,2)) + c(-1)^{1+3}\det(A(1,3)) \\
&= a(-1)^2\det\left(\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) + b(-1)^3\det\left(\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c(-1)^4\det\left(\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right) \\
&= a(ei - fh) - b(di - fg) + c(dh - eg) \\
&= aei - afh - bdi + bfg + cdh - ceg.
\end{aligned}$$

98. Performing a cofactor expansion along the first column of B gives

$$\begin{aligned}
\det(B) = \begin{vmatrix} 1 & 0 & -2 \\ 0 & 3 & 4 \\ 3 & 6 & 2 \end{vmatrix} &= 1\begin{vmatrix} 3 & 4 \\ 6 & 2 \end{vmatrix} - 0\begin{vmatrix} 0 & -2 \\ 6 & 2 \end{vmatrix} + 3\begin{vmatrix} 0 & -2 \\ 3 & 4 \end{vmatrix} \\
&= 1(6-24) + 0 + 3(0+6) \\
&= -18 + 18 \\
&= 0.
\end{aligned}$$

Since $\det(B) = 0$, B is not invertible.

99. If the ith row of $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ is a row of zeros, then $a_{ij} = 0$ for $j = 1, \ldots, n$ Performing a cofactor expansion along the ith row of A gives

$$\det(A) = a_{i1}C_{i1}(A) + a_{i2}C_{i2}(A) + \cdots + a_{in}C_{in}(A) = 0C_{i1}(A) + 0C_{i2}(A) + \cdots + 0C_{in}(A) = 0.$$

If the jth column of $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ is a column of zeros, then $a_{ij} = 0$ for $i = 1, \ldots, n$. Performing a cofactor expansion along the jth column of A gives

$$\det(A) = a_{1j}C_{1j}(A) + a_{2j}C_{2,j}(A) + \cdots + a_{nj}C_{nj}(A) = 0C_{1j}(A) + 0C_{2,j}(A) + \cdots + 0C_{nj}(A).$$

6.2 Elementary Row and Column Operations

100. Note that

$$\det(A) = 12, \quad \det(B) = -12, \quad \det(C) = 12 \quad \text{and} \quad \det(D) = -36.$$

Now

$$A = \begin{bmatrix} 2 & -1 \\ 6 & 3 \end{bmatrix} \underset{R_1 \leftrightarrow R_2}{\longrightarrow} \begin{bmatrix} 6 & 3 \\ 2 & -1 \end{bmatrix} = B \quad \text{and} \quad \det(B) = -\det(A)$$
$$A = \begin{bmatrix} 2 & -1 \\ 6 & 3 \end{bmatrix} \underset{R_2 - 2R_1}{\longrightarrow} \begin{bmatrix} 2 & -1 \\ 2 & 5 \end{bmatrix} = C \quad \text{and} \quad \det(C) = \det(A)$$
$$A = \begin{bmatrix} 2 & -1 \\ 6 & 3 \end{bmatrix} \overset{-3R_1}{\longrightarrow} \begin{bmatrix} -6 & 3 \\ 6 & 3 \end{bmatrix} = D \quad \text{and} \quad \det(D) = -3\det(A)$$

101. We have

$$\begin{aligned}
\det(A) = \begin{vmatrix} x & x & 1 \\ x & 1 & x \\ 1 & x & x \end{vmatrix} \overset{\substack{R_1 - xR_3 \\ R_2 - xR_3}}{=} \begin{vmatrix} 0 & x - x^2 & 1 - x^2 \\ 0 & 1 - x^2 & x - x^2 \\ 1 & x & x \end{vmatrix} \\
&= 1 \begin{vmatrix} x(1-x) & (1+x)(1-x) \\ (1+x)(1-x) & x(1-x) \end{vmatrix} \\
&= (1-x)^2 \begin{vmatrix} x & 1+x \\ 1+x & x \end{vmatrix} \\
&= (1-x)^2(x^2 - (1+x)^2) \\
&= (1-x)^2(x^2 - 1 - 2x - x^2) \\
&= -(1-x)^2(1+2x)
\end{aligned}$$

Now A fails to be invertible exactly when $\det(A) = 0$, that is, when $-(1-x)^2(1+2x) = 0$. Thus we have $x = 1$ or $x = -1/2$.

102. We have

$$\det(A) = \begin{bmatrix} -1 & 4 & 3 \\ 2 & 0 & -2 \\ 2 & 3 & -2 \end{bmatrix} \underset{\substack{C_2 + 4C_1 \to C_2 \\ C_3 + 3C_1 \to C_3}}{=} \begin{bmatrix} -1 & 0 & 0 \\ 2 & 8 & 4 \\ 2 & 11 & 4 \end{bmatrix} \underset{C_3 - \frac{1}{2}C_2 \to C_3}{=} \begin{bmatrix} -1 & 0 & 0 \\ 2 & 8 & 0 \\ 2 & 11 & -3/2 \end{bmatrix},$$

so

$$\det(A) = -1(8)\left(-\frac{3}{2}\right) = 12.$$

6.3 Properties of Determinants

103. Since $\det(-2A) = (-2)^n \det(A)$ by Theorem 6.3.2, we see that $(-2)^n = 64$. Since $64 = (-2)^6$, it follows that $n = 6$.

104.

(a) Using elementary row operations, we have

$$\begin{aligned}\det(A) = \begin{vmatrix} 1 & 1 & 2 \\ -1 & 3 & 0 \\ 1 & 2 & 1 \end{vmatrix} \underset{\substack{R_2+R_1 \\ R_3-R_1}}{=} \begin{vmatrix} 1 & 1 & 2 \\ 0 & 4 & 2 \\ 0 & 1 & -1 \end{vmatrix} \underset{R_3-\frac{1}{4}R_2}{=} \begin{vmatrix} 1 & 1 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & -3/2 \end{vmatrix} \\ = 1(4)\left(-\frac{3}{2}\right) = -6.\end{aligned}$$

(b) Using elementary column operations, we have

$$\begin{aligned}\det(A^T) = \begin{vmatrix} 1 & -1 & 1 \\ 1 & 3 & 2 \\ 2 & 0 & 1 \end{vmatrix} \underset{\substack{C_2+C_1\to C_2 \\ C_3-C_1\to C_3}}{=} \begin{vmatrix} 1 & 0 & 0 \\ 1 & 4 & 1 \\ 2 & 2 & -1 \end{vmatrix} \underset{C_3-\frac{1}{4}C_2\to C_3}{=} \begin{vmatrix} 1 & 0 & 0 \\ 1 & 4 & 0 \\ 2 & 2 & -3/2 \end{vmatrix} \\ = 1(4)\left(-\frac{3}{2}\right) = -6.\end{aligned}$$

(c) The column operations in part (b) correspond to the row operations in part (a). That is, if we use a sequence of elementary row operations to carry A to an upper triangular form when computing $\det(A)$, we can perform the sequence of corresponding elementary column operations to A^T to carry A^T to a lower triangular form when computing $\det(A^T)$. The resulting diagonal entries from either case will be the same, so the determinants will be the same.

105. Consider $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Then

$$\det(A) = \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = 1 - 0 = 1 \quad \text{and} \quad \det(B) = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 - 0 = 1$$

so $\det(A) + \det(B) = 1 + 1 = 2$. Now

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

Thus $\det(A + B) = \det(A) + \det(B)$ in this case.

6.4 Optional Section: Area, Volume and Determinants

106. From the table at the end of Section 5.2, we have that

$$\big[\,T\,\big] = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}.$$

Thus by Theorem 6.4.3,

$$\operatorname{area}(T(P)) = \left|\det\left(\big[\,T\,\big]\right)\right| \operatorname{area}(P) = |1|(a) = a.$$

We see the area of P does not change under the transformation T.

107. By the table at the end of Section 5.2,

$$\left[T\right] = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}.$$

It follows that

$$\det\left(\left[T\right]\right) = \det\left(\begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\right) \det\left(\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\right) = \frac{1}{4}(1) = \frac{1}{4}.$$

Thus, by Theorem 6.4.3,

$$\text{area}(T(Q)) = \left|\det\left(\left[T\right]\right)\right| \text{area}(Q) = \left|\frac{1}{4}\right| (2) = \frac{1}{2}.$$

108. Since

$$\begin{aligned} \det\left(\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right) &= \left(\begin{bmatrix} -1 & 1 & 3 \\ -2 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix}\right) \\ &= 3\begin{vmatrix} -2 & 1 \\ 1 & 1 \end{vmatrix} + 2\begin{vmatrix} -1 & 1 \\ -2 & 1 \end{vmatrix} \\ &= 3(-2-1) + 2(-1+2) \\ &= -7, \end{aligned}$$

Theorem 6.4.6 gives that

$$\text{vol}(Q) = \left|\det\left(\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right)\right| = |-7| = 7.$$

109. We have

$$\begin{aligned} \text{vol}(T(Q)) &= \left|\det\left(\begin{bmatrix} T(\vec{x}) & T(\vec{y}) & T(\vec{z}) \end{bmatrix}\right)\right| && \text{by Theorem 6.4.6} \\ &= \left|\det\left(\begin{bmatrix} [T]\,\vec{x} & [T]\,\vec{y} & [T]\,\vec{z} \end{bmatrix}\right)\right| \\ &= \left|\det\left([T]\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right)\right| && \text{by Definition 3.4.1} \\ &= \left|\det\left([T]\right)\det\left(\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right)\right| && \text{by Theorem 6.3.4} \\ &= \left|\det\left([T]\right)\right| \left|\det\left(\begin{bmatrix} \vec{x} & \vec{y} & \vec{z} \end{bmatrix}\right)\right| \\ &= \left|\det\left([T]\right)\right| \text{vol}(Q) && \text{by Theorem 6.4.6.} \end{aligned}$$

6.5 Optional Section: Adjugates and Matrix Inverses

110.

(a) Let

$$D = AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} C_{11}(A) & C_{21}(A) & C_{31}(A) \\ C_{12}(A) & C_{22}(A) & C_{32}(A) \\ C_{13}(A) & C_{23}(A) & C_{33}(A) \end{bmatrix}.$$

We see that the diagonal entries of D are

$$\begin{aligned} d_{11} &= a_{11}C_{11}(A) + a_{12}C_{12}(A) + a_{13}C_{13}(A) \\ d_{22} &= a_{21}C_{21}(A) + a_{22}C_{22}(A) + a_{23}C_{23}(A) \end{aligned}$$

$$d_{33} = a_{31}C_{31}(A) + a_{32}C_{32}(A) + a_{33}C_{33}(A)$$

which we recognize as cofactor expansions along the first, second and third rows of A, respectively. Thus $d_{11} = d_{22} = d_{33} = \det(A)$. We now compute the $(1,2)-$ and $(1,3)-$ entries of D.

$$\begin{aligned}
d_{12} &= a_{11}C_{21}(A) + a_{12}C_{22}(A) + a_{13}C_{23}(A) \\
&= a_{11}\big(-(a_{12}a_{33} - a_{13}a_{32})\big) + a_{12}(a_{11}a_{33} - a_{13}a_{31}) + a_{13}\big(-(a_{11}a_{32} - a_{12}a_{31})\big) \\
&= -a_{11}a_{12}a_{33} + a_{11}a_{13}a_{32} + a_{11}a_{12}a_{33} - a_{12}a_{13}a_{31} - a_{11}a_{13}a_{32} + a_{12}a_{13}a_{31} \\
&= 0 \\
d_{13} &= a_{11}C_{31}(A) + a_{12}C_{32}(A) + a_{13}C_{33}(A) \\
&= a_{11}(a_{12}a_{23} - a_{13}a_{22}) + a_{12}\big(-(a_{11}a_{23} - a_{13}a_{21})\big) + a_{13}(a_{11}a_{22} - a_{12}a_{21}) \\
&= a_{11}a_{12}a_{23} - a_{11}a_{13}a_{22} - a_{11}a_{12}a_{23} + a_{12}a_{13}a_{21} + a_{11}a_{13}a_{22} - a_{12}a_{13}a_{21} \\
&= 0
\end{aligned}$$

We can show that d_{21}, d_{23}, d_{31} and d_{32} are all zero in a similarly tedious fashion. Thus $D = AB = \det(A)I_3$.

(b) If $\det(A) \neq 0$, then from $AB = \det(A)I_3$ we have

$$A\left(\frac{1}{\det(A)}B\right) = I_3$$

so

$$A^{-1} = \frac{1}{\det(A)}B.$$

111.

(a) Since

$$\det(A) = 9(-5) - 7(-7) = -45 + 49 = 4$$

and

$$\operatorname{adj}(A) = \begin{bmatrix} -5 & 7 \\ -7 & 9 \end{bmatrix},$$

we have

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A) = \frac{1}{4}\begin{bmatrix} -5 & 7 \\ -7 & 9 \end{bmatrix} = \begin{bmatrix} -5/4 & 7/4 \\ -7/4 & 9/4 \end{bmatrix}.$$

(b) We have

$$\left[\begin{array}{cc|cc} 9 & -7 & 1 & 0 \\ 7 & -5 & 0 & 1 \end{array}\right] \xrightarrow{\frac{1}{9}R_1} \left[\begin{array}{cc|cc} 1 & -7/9 & 1/9 & 0 \\ 7 & -5 & 0 & 1 \end{array}\right] \xrightarrow[R_2-7R_1]{} \left[\begin{array}{cc|cc} 1 & -7/9 & 1/9 & 0 \\ 0 & 4/9 & -7/9 & 1 \end{array}\right] \xrightarrow[\frac{9}{4}R_2]{}$$
$$\left[\begin{array}{cc|cc} 1 & -7/9 & 1/9 & 0 \\ 0 & 1 & -7/4 & 9/4 \end{array}\right] \xrightarrow{R_1+\frac{7}{9}R_2} \left[\begin{array}{cc|cc} 1 & 0 & 5/4 & 7/4 \\ 0 & 1 & -7/4 & 9/4 \end{array}\right]$$

so

$$A^{-1} = \begin{bmatrix} -5/4 & 7/4 \\ -7/4 & 9/4 \end{bmatrix}.$$

112.

(a) Note that

$$\operatorname{adj}(A) = \begin{bmatrix} -3 & 2 & 1 \\ 1 & -4 & 1 \\ 1 & 2 & -1 \end{bmatrix}$$

by Example 6.5.4. Since $A \operatorname{adj}(A) = \det(A) I_3$ by Theorem 6.5.5, we can compute $\det(A)$ by computing the $(1,1)-$entry of $A \operatorname{adj}(A)$. We have

$$\det(A) = 1(-3) + 2(1) + 3(1) = 2.$$

Thus

$$A^{-1} = \frac{1}{\det(A)} = \frac{1}{2} \begin{bmatrix} -3 & 2 & 1 \\ 1 & -4 & 1 \\ 1 & 2 & -1 \end{bmatrix} = \begin{bmatrix} -3/2 & 1 & 1/2 \\ 1/2 & -2 & 1/2 \\ 1/2 & 1 & -1/2 \end{bmatrix}.$$

(b) We have

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 3 & 4 & 5 & 0 & 0 & 1 \end{array}\right] \underset{\substack{R_2 - R_1 \\ R_3 - 3R_1}}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -1 & -1 & 1 & 0 \\ 0 & -2 & -4 & -3 & 0 & 1 \end{array}\right] \underset{R_3 - 2R_2}{\overset{R_1 + 2R_2}{\longrightarrow}}$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & -1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 & 0 \\ 0 & 0 & -2 & -1 & -2 & 1 \end{array}\right] \underset{\substack{R_2 \\ -\frac{1}{2}R_3}}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & -1 & 2 & 0 \\ 0 & 1 & 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1/2 & 1 & -1/2 \end{array}\right] \underset{\substack{R_1 - R_3 \\ R_2 - R_3}}{\longrightarrow}$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -3/2 & 1 & 1/2 \\ 0 & 1 & 0 & 1/2 & -2 & 1/2 \\ 0 & 0 & 1 & 1/2 & 1 & -1/2 \end{array}\right].$$

Thus

$$A^{-1} = \begin{bmatrix} -3/2 & 1 & 1/2 \\ 1/2 & -2 & 1/2 \\ 1/2 & 1 & -1/2 \end{bmatrix}.$$

7.1 Basic Operations

113. For $x, y \in \mathbb{R}$, we have

$$(x + 0i) + (y + 0i) = (x + y) + (0 + 0)i = x + y$$
$$(x + 0i)(y + 0i) = xy + x0i + 0yi + 0^2 i^2 = xy$$

which shows that our definitions of addition and multiplication of complex numbers are consistent with addition and multiplication of real numbers.

114. We compute

$$\begin{aligned} zw &= (x + yi)\left(\frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2} i\right) \\ &= \frac{x^2}{x^2 + y^2} - \frac{xy}{x^2 + y^2} i + \frac{yx}{x^2 + y^2} i - \frac{y^2}{x^2 + y^2} i^2 \\ &= \frac{x^2}{x^2 + y^2} + \frac{y^2}{x^2 + y^2} \end{aligned}$$

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

115. We carry out our operations as we would with real numbers.

$$\begin{aligned} \frac{(1-2i)-(3+4i)}{5-6i} &= \frac{-2-6i}{5-6i} \\ &= \frac{-2-6i}{5-6i}\left(\frac{5+6i}{5+6i}\right) \\ &= \frac{-10-12i-30i-36i^2}{25+36} \\ &= \frac{26-42i}{61} \\ &= \frac{26}{61}-\frac{42}{61}i \end{aligned}$$

7.3 Polar Form

116. Recall that

$$\begin{aligned} \cos(\theta_1-\theta_2) &= \cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2 \\ \sin(\theta_1-\theta_2) &= \sin\theta_1\cos\theta_2-\cos\theta_1\sin\theta_2. \end{aligned}$$

So

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{r_1(\cos\theta_1+i\sin\theta_1)}{r_2(\cos\theta_2+i\sin\theta_2)} \\ &= \frac{r_1}{r_2}\,\frac{\cos\theta_1+i\sin\theta_1}{\cos\theta_2+i\sin\theta_2}\,\frac{\cos\theta_2-i\sin\theta_2}{\cos\theta_2-i\sin\theta_2} \\ &= \frac{r_1}{r_2}\,\frac{(\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2)+i(\sin\theta_1\cos\theta_2-\cos\theta_1\sin\theta_2)}{\cos^2\theta_2+\sin^2\theta_2} \\ &= \frac{r_1}{r_2}\big(\cos(\theta_1-\theta_2)+i\sin(\theta_1-\theta_2)\big). \end{aligned}$$

117. Note that $|1|=1$ and $\theta=0$ is an argument for 1. Using the result of Exercise 116, we have

$$z^{-1}=\frac{1}{z}=\frac{1(\cos 0+i\sin 0)}{r(\cos\theta+i\sin\theta)}=\frac{1}{r}\big(\cos(0-\theta)+i\sin(0-\theta)\big)=\frac{1}{r}\big(\cos(-\theta)+i\sin(-\theta)\big).$$

118. Since $r=\left|\frac{1}{2}+\frac{\sqrt{3}}{2}i\right|=\sqrt{\frac{1}{4}+\frac{3}{4}}=1$, we see that

$$\frac{1}{2}+\frac{\sqrt{3}}{2}i=\cos\frac{\pi}{3}+i\sin\frac{\pi}{3}.$$

Thus

$$\left(\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)^{602} = \left(\cos\frac{\pi}{3}+i\sin\frac{\pi}{3}\right)^{602}$$

$$\begin{aligned}
&= \cos\frac{602\pi}{3} + i\sin\frac{602\pi}{3} \quad \text{by de Moivre's Theorem}\\
&= \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3}\\
&= -\frac{1}{2} + \frac{\sqrt{3}}{2}\,i.
\end{aligned}$$

119. For $z = 1$, we have $|z| = |1| = 1$ and that $\theta = 0$ is an argument for z. Thus

$$z = 1 = 1\,(\cos(0) + i\sin(0)) = e^{i0}.$$

For $w = i$, we have $|w| = |i| = 1$ and that $\theta = \pi/2$ is an argument for w. Thus

$$w = i = 1\,(\cos(\pi/2) + i\sin(\pi/2)) = e^{i\frac{\pi}{2}}.$$

120. Let $z_1 = r_1e^{i\theta_1}$ and $z_2 = r_2e^{i\theta_2}$. Then

$$\begin{aligned}
z_1z_2 &= r_1e^{i\theta_1}r_2e^{i\theta_2}\\
&= r_1(\cos\theta_1 + i\sin\theta_1)r_2(\cos\theta_2 + i\sin\theta_2)\\
&= r_1r_2(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)) \qquad \text{by Theorem 7.3.4}\\
&= r_1r_2e^{i(\theta_1+\theta_2)}.
\end{aligned}$$

7.4 Complex Polynomials

121.

(a) Adding corresponding coefficients, we have

$$\begin{aligned}
p(z)+q(z) &= \big((1+i)+0\big)z^4 + (0+5)z^3 + \big(-(2-i)+(2+i)\big)z^2 + (4i+0)z+\\
&\qquad + (4-2-i)\\
&= (1+i)z^4 + 5z^3 + 2iz^2 + 4iz + 2 - i.
\end{aligned}$$

(b) Multiplying the coefficients of $q(z)$ by i gives

$$\begin{aligned}
iq(z) &= i\big(5z^3 + (2+i)z^2 - 2 - i\big)\\
&= 5iz^3 + (-1+2i)z^2 + 1 - 2i.
\end{aligned}$$

7.5 Complex nth Roots

122. Since $|4-4\sqrt{3}i| = 4|1-\sqrt{3}i| = 4\sqrt{1+3} = 4(2) = 8$, we have

$$4-4\sqrt{3}i = 8\left(\frac{4}{8} - \frac{4\sqrt{3}}{8}\,i\right) = 8\left(\frac{1}{2} - \frac{\sqrt{3}}{2}\,i\right) = 8\left(\cos\frac{5\pi}{3} + i\sin\frac{5\pi}{3}\right).$$

Thus, the 3rd roots of $4-4\sqrt{3}i$ are given by

$$z_k = 8^{1/3}\left(\cos\left(\frac{\frac{5\pi}{3}+2k\pi}{3}\right) + i\sin\left(\frac{\frac{5\pi}{3}+2k\pi}{3}\right)\right)$$

$$= 2\left(\cos\left(\frac{5\pi + 6k\pi}{9}\right) + i\sin\left(\frac{5\pi + 6k\pi}{9}\right)\right)$$

for $k = 0, 1, 2$. These are:

$$\begin{aligned}
z_0 &= 2\left(\cos\frac{5\pi}{9} + i\sin\frac{5\pi}{9}\right)\\
z_1 &= 2\left(\cos\frac{11\pi}{9} + i\sin\frac{11\pi}{9}\right)\\
z_2 &= 2\left(\cos\frac{17\pi}{9} + i\sin\frac{17\pi}{9}\right).
\end{aligned}$$

8.1 Introduction

123.

(a) For $\vec{x} \in L$, there is a $t \in \mathbb{R}$ such that $\vec{x} = t\vec{d}$. Thus

$$\begin{aligned}
T(\vec{x}) &= T(t\vec{d})\\
&= 2\operatorname{proj}_{\vec{d}}(t\vec{d}) - t\vec{d}\\
&= 2\frac{(t\vec{d})\cdot\vec{d}}{\|\vec{d}\|^2}\vec{d} - t\vec{d}\\
&= 2t\frac{\vec{d}\cdot\vec{d}}{\|\vec{d}\|^2}\vec{d} - t\vec{d}\\
&= 2t\vec{d} - t\vec{d} && \text{since } \vec{d}\cdot\vec{d} = \|\vec{d}\|^2 \neq 0\\
&= t\vec{d}\\
&= \vec{x}.
\end{aligned}$$

(b) For $\vec{x} \in L'$, there is an $s \in \mathbb{R}$ such that $\vec{x} = s\vec{n}$ where $\vec{n}$ is any direction vector for L'. Since L and L' are perpendicular, we have that $\vec{d}\cdot\vec{n} = 0$. It follows that

$$\begin{aligned}
T(\vec{x}) &= T(s\vec{n})\\
&= 2\operatorname{proj}_{\vec{d}}(s\vec{n}) - s\vec{n}\\
&= 2\frac{(s\vec{n})\cdot\vec{d}}{\|\vec{d}\|^2}\vec{d} - s\vec{n}\\
&= 2s\frac{\vec{n}\cdot\vec{d}}{\|\vec{d}\|^2}\vec{d} - s\vec{n}\\
&= 2s(0)\vec{d} - s\vec{n} && \text{since } \vec{n}\cdot\vec{d} = 0\\
&= -s\vec{n}\\
&= -\vec{x}.
\end{aligned}$$

124.

(a) We begin by noting that since $\vec{n} = \left[\begin{smallmatrix}1\\2\\3\end{smallmatrix}\right]$ is a normal vector for P, we have that $\vec{v}_1 \cdot \vec{n} = 0 = \vec{v}_2 \cdot \vec{n}$. Thus

$$T(\vec{v}_1) = \vec{v}_1 - \operatorname{proj}_{\vec{n}} \vec{v}_1 = \vec{v}_1 - \frac{\vec{v}_1 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \vec{v}_1 = 1\vec{v}_1,$$
$$T(\vec{v}_2) = \vec{v}_2 - \operatorname{proj}_{\vec{n}} \vec{v}_2 = \vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \vec{v}_2 = 1\vec{v}_2,$$

which shows that both $\vec{v}_1, \vec{v}_2$ are eigenvectors of T corresponding to $\mu_1 = 1$.

(b) Let $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2$ be nonzero for some $c_1, c_2 \in \mathbb{R}$. Then

$$T(\vec{x}) = T(c_1\vec{v}_1 + c_2\vec{v}_2) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) = c_1\vec{v}_1 + c_1\vec{v}_2 = 1\vec{x}$$

so any nonzero linear combination of $\vec{v}_1$ and $\vec{v}_2$ is also an eigenvector of T corresponding to $\mu_1 = 1$.

(c) We have

$$T(\vec{n}) = \vec{n} - \operatorname{proj}_{\vec{n}} \vec{n} = \vec{n} - \vec{n} = \vec{0} = 0\vec{n}$$

which shows that $\vec{n}$ is an eigenvector of T corresponding to $\mu_2 = 0$.

(d) Let $\vec{x} = c\vec{n}$ be nonzero for some $c \in \mathbb{R}$. Then

$$T(\vec{x}) = T(c\vec{n}) = cT(\vec{n}) = c(\vec{0}) = \vec{0} = 0\vec{x}$$

so any nonzero linear combination of $\vec{n}$ is also an eigenvector of T corresponding to $\mu_2 = 0$.

125. There are infinitely many bases for P. To find one such basis, let $\vec{x} = \left[\begin{smallmatrix}x_1\\x_2\\x_3\end{smallmatrix}\right] \in P$. Then $x_1 + 2x_2 + x_3 = 0$ so $x_3 = -x_1 - 2x_2$. We have

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

Letting

$$\vec{v}_1 = \begin{bmatrix}1\\0\\-1\end{bmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{bmatrix}0\\1\\-2\end{bmatrix},$$

we see that $\operatorname{Span} B = P$. Since neither vector in B is a scalar multiple of the other, B is linearly independent and thus a basis for P.

8.2 Computing Eigenvalues and Eigenvectors

126. Since $A\vec{x} = 1\vec{x}$ for all $\vec{x} \in \mathbb{R}^2$, we see that $\lambda = 1$ is an eigenvalue of A and every nonzero $\vec{x} \in \mathbb{R}^2$ is an eigenvector.

127. We have

$$\det(A-\lambda I)=\det\left(\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix}-\lambda\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right)=\begin{vmatrix}1-\lambda&0&1\\0&1-\lambda&0\\1&0&1-\lambda\end{vmatrix}.$$

Performing a cofactor expansion along the second row of $A-\lambda I$ gives

$$\begin{aligned}\det(A-\lambda I)&=(1-\lambda)\begin{vmatrix}1-\lambda&1\\1&1-\lambda\end{vmatrix}\\&=(1-\lambda)\big((1-\lambda)^2-1\big)\\&=(1-\lambda)(\lambda^2-2\lambda)\\&=-\lambda(\lambda-1)(\lambda-2),\end{aligned}$$

from which we see that $\det(A-\lambda I)=0$ if and only if $\lambda=0$, $\lambda=1$ or $\lambda=2$. Thus the eigenvalues of A are $\lambda_1=0$, $\lambda_2=1$ and $\lambda_3=2$.

128. From Exercise 127, we know that $\lambda_1=0$, $\lambda_2=1$ and $\lambda_3=2$ are the eigenvalues of A. For $\lambda_1=0$, we have

$$A-\lambda_1 I=A=\begin{bmatrix}1&0&1\\0&1&0\\1&0&1\end{bmatrix}\underset{R_3-R_1}{\longrightarrow}\begin{bmatrix}1&0&1\\0&1&0\\0&0&0\end{bmatrix}$$

so the eigenvectors of A corresponding to $\lambda_1=0$ are

$$\vec{x}=\begin{bmatrix}-t\\0\\t\end{bmatrix}=t\begin{bmatrix}-1\\0\\1\end{bmatrix},\quad t\in\mathbb{R},\,t\neq 0.$$

For $\lambda_2=1$, we have

$$A-\lambda_2 I=A-I=\begin{bmatrix}0&0&1\\0&0&0\\1&0&0\end{bmatrix}\underset{R_1\leftrightarrow R_3}{\longrightarrow}\begin{bmatrix}1&0&0\\0&0&0\\0&0&1\end{bmatrix}\underset{R_2\leftrightarrow R_3}{\longrightarrow}\begin{bmatrix}1&0&0\\0&0&1\\0&0&0\end{bmatrix}$$

so the eigenvectors of A corresponding to $\lambda_2=1$ are

$$\vec{x}=\begin{bmatrix}0\\t\\0\end{bmatrix}=t\begin{bmatrix}0\\1\\0\end{bmatrix},\quad t\in\mathbb{R},\,t\neq 0.$$

Finally, for $\lambda_3=2$, we have

$$A-\lambda_3 I=A-2I=\begin{bmatrix}-1&0&1\\0&-1&0\\1&0&-1\end{bmatrix}\underset{\longrightarrow}{\substack{-R_1\\-R_2}}\begin{bmatrix}1&0&-1\\0&1&0\\1&0&-1\end{bmatrix}\underset{R_3-R_1}{\longrightarrow}\begin{bmatrix}1&0&-1\\0&1&0\\0&0&0\end{bmatrix}$$

from which we see that the eigenvectors of A corresponding to $\lambda_3=2$ are

$$\vec{x}=\begin{bmatrix}t\\0\\t\end{bmatrix}=t\begin{bmatrix}1\\0\\1\end{bmatrix},\quad t\in\mathbb{R},\,t\neq 0.$$

129. Notice that A is upper (and lower) triangular. By Theorem 8.2.9, the only eigenvalue of A is $\lambda=1$. We have

$$A-I=\begin{bmatrix}0&0\\0&0\end{bmatrix}$$

so the eigenvalues of A corresponding to $\lambda = 1$ are

$$\vec{x} = \begin{bmatrix} s \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad s, t \in \mathbb{R},\ s, t \text{ not both } 0.$$

130. Since a rotation does not change the norm of a vector, we conclude that if $A(\vec{x}) = \lambda\vec{x}$, then $\lambda = \pm 1$. If $\lambda = 1$, then $\theta = 0$ and if $\lambda = -1$, then $\theta = \pi$.

Alternatively, we compute

$$\begin{aligned} C_A(\lambda) = \begin{vmatrix} \cos\theta - \lambda & -\sin\theta \\ \sin\theta & \cos\theta - \lambda \end{vmatrix} &= (\cos\theta - \lambda)^2 + \sin^2\theta \\ &= \cos^2\theta - 2\lambda\cos\theta + \lambda^2 + \sin^2\theta \\ &= \lambda^2 - 2\lambda\cos\theta + 1. \end{aligned}$$

In order to have $\lambda \in \mathbb{R}$, we require the discriminant of $C_A(\lambda)$ to be nonnegative. Thus, we require

$$(-2\cos\theta)^2 - 4(1)(1) \geq 0.$$

From this, we see that we must have $\cos^2\theta = 1$. This occurs exactly when $\cos\theta = \pm 1$. Hence we must have $\theta = 0, \pi$.

8.3 Eigenspaces

131. We know from Exercise 128 that $\lambda_1 = 0$, $\lambda_2 = 1$ and $\lambda_3 = 2$ are the eigenvalues of A. We also know that the solutions to $A\vec{x} = \vec{0}$ are

$$\vec{x} = t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

so

$$B_1 = \left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_1}(A)$ and $\dim(E_{\lambda_1}(A)) = 1$. We also computed that

$$\vec{x} = t\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}$$

is a solution to $(A - I)\vec{x} = \vec{0}$, so

$$B_2 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_2}(A)$ and $\dim(E_{\lambda_2}(A)) = 1$. Finally, we derived that the solution to $(A - 2I)\vec{x} = \vec{0}$ is

$$\vec{x} = t\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}$$

giving that

$$B_3 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_3}(A)$ and $\dim(E_{\lambda_3}(A)) = 1$.

132. Since A is upper triangluar, the characteristic polynomial of A is

$$C_A(\lambda) = (\lambda + 4)(\lambda - 3)^2.$$

Thus the eigenvalues are $\lambda_1 = -4$ with algebraic multiplicity $a_{\lambda_1} = 1$ and $\lambda_2 = 3$ with algebraic multiplicity $a_{\lambda_2} = 2$.

133. From Exercise 132, $\lambda_1 = -4$ and $\lambda_2 = 3$. We find a basis for each eigenspace of A. For $\lambda_1 = -4$, we solve $(A + 4I)\vec{x} = \vec{0}$. Since

$$A + 4I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 7 & 1 \\ 0 & 0 & 7 \end{bmatrix} \underset{\substack{\frac{1}{7}R_2 \\ \frac{1}{7}R_3}}{\longrightarrow} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1/7 \\ 0 & 0 & 1 \end{bmatrix} \underset{R_2 - \frac{1}{7}R_3}{\longrightarrow} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

we have that

$$\vec{x} = \begin{bmatrix} t \\ 0 \\ 0 \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}$$

so

$$B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_1}(A)$ and $g_{\lambda_1} = \dim(E_{\lambda_1}(A)) = 1$. For $\lambda_2 = 3$, we solve $(A - 3I)\vec{x} = \vec{0}$. Since

$$A - 3I = \begin{bmatrix} -7 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \overset{-\frac{1}{7}R_1}{\longrightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

we have that

$$\vec{x} = \begin{bmatrix} 0 \\ t \\ 0 \end{bmatrix} = t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad t \in \mathbb{R}$$

so

$$B_2 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis for $E_{\lambda_2}(A)$ and $g_{\lambda_2} = \dim(E_{\lambda_2}(A)) = 2$.

8.4 Diagonalization

134. We consider the matrix

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Since

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \underset{R_3 - R_1}{\longrightarrow} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \underset{R_2 \leftrightarrow R_3}{\longrightarrow} \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

we see that $\operatorname{rank}(P) = 3$ so B is a basis (consisting of eigenvectors of A) for $\mathbb{R}^3$ by Theorem 4.5.7.

135. From Example 8.3.9, the eigenvalues of A are $\lambda_1 = -1$ and $\lambda_2 = 2$ with $a_{\lambda_1} = 2$ and $a_{\lambda_2} = 1$. Additionally,

$$B_1 = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad B_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

are bases for $E_{\lambda_1}(A)$ and $E_{\lambda_2}(A)$ respectively, so $g_{\lambda_1} = 2$ and $g_{\lambda_2} = 1$. Since $a_{\lambda_1} = g_{\lambda_1}$ and $a_{\lambda_2} = g_{\lambda_2}$, we see that A is diagonalizable. We let

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

so that

$$P^{-1}AP = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = D.$$

136. The set

$$B = B_1 \cup B_2 = \left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \end{bmatrix} \right\}$$

is linearly independent by Theorem 8.4.6. Since B contains 2 vectors, it follows from Corollary 4.5.9 that B spans $\mathbb{R}^2$ and is thus a basis for $\mathbb{R}^2$. Since B consists of eigenvectors of A, we have that A is diagonalizable by the Diagonalization Theorem. We let

$$P = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$$

and compute P^{-1} using the Matrix Inversion Algorithm.

$$\left[\begin{array}{cc|cc} 1 & -2 & 1 & 0 \\ 3 & 2 & 0 & 1 \end{array}\right] \underset{R_2 - 3R_1}{\longrightarrow} \left[\begin{array}{cc|cc} 1 & -2 & 1 & 0 \\ 0 & 8 & -3 & 1 \end{array}\right] \underset{\frac{1}{8}R_2}{\longrightarrow}$$
$$\left[\begin{array}{cc|cc} 1 & -2 & 1 & 0 \\ 0 & 1 & -3/8 & 1/8 \end{array}\right] \overset{R_1 + 2R_2}{\longrightarrow} \left[\begin{array}{cc|cc} 1 & 0 & 1/4 & 1/4 \\ 0 & 1 & -3/8 & 1/8 \end{array}\right],$$

so

$$P^{-1} = \begin{bmatrix} 1/4 & 1/4 \\ -3/8 & 1/8 \end{bmatrix}.$$

Thus

$$\begin{aligned} A = PDP^{-1} &= \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1/4 & 1/4 \\ -3/8 & 1/8 \end{bmatrix} \\ &= \begin{bmatrix} 2 & -12 \\ 6 & 12 \end{bmatrix} \left(\frac{1}{8} \begin{bmatrix} 2 & 2 \\ -3 & 1 \end{bmatrix} \right) \\ &= \frac{1}{8} \begin{bmatrix} 40 & -8 \\ -24 & 24 \end{bmatrix} \\ &= \begin{bmatrix} 5 & -1 \\ -3 & 3 \end{bmatrix}. \end{aligned}$$