Linear Algebra 1

Course Notes for MATH 136

Edition 1.0

D. Wolczuk

MEF
Mathematics Endowment Fund

Contents

A NOTE TO STUDENTS - READ THIS!

Best Selling Novel: "My Favourite Math," by Al G. Braw

Welcome to Linear Algebra!

In this course, you will be introduced to basic elements in linear algebra, including vectors, vector spaces, matrices, linear mappings, and solving systems of linear equations. The strategy used in this course is to help you understand an example of a concept before generalizing it. For example, in Chapter 1, you will learn about vectors in $\mathbb{R}^n$, and then in Chapter 4, we will extend part of what we did with vectors in $\mathbb{R}^n$ to the abstract setting of a vector space. Of course, the better you understand the particular example, the easier it will be for you to understand the more complicated concepts introduced later.

As you will quickly see, this course contains both computational and theoretical elements. Most students, with enough practice, will find the computational questions (about 70% of the tests) fairly easy, but may have difficulty with the theory portion (definitions, theorems and proofs). While studying this course, you should keep in mind that when applying the material in the future, problems will likely be far too large to compute by hand. Hence, the computational work will be done by a computer, and you will use your knowledge and understanding of the *theory* to interpret the results. Why do we make you compute things by hand if they can be done by a computer? Because solving problems by hand can often help you understand the theory. Thus, when you first learn how to solve problems in this course, pay close attention to how solving the problems applies to the concepts in the course (definitions and theorems). In doing so, not only will you find the computational problems easier, but it will also help you greatly when doing proofs.

Prerequisites:

Teacher: Recall from your last course that...
Student: What?!? You mean we were supposed to remember that?

We expect that you know:

- How to solve a system of equations using substitution and elimination
- The basics of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, including the dot product and cross product
- The basics of how to write proofs
- Basic operations on complex numbers

If you are unsure of any of these, we recommend that you do some self study to learn/remember these topics.

What is linear algebra used for?

> *Student: Will we ever use this in real life?*
>
> *Professor: Not if you get a job flipping hamburgers.*

Linear algebra is used in the social sciences, the natural sciences, engineering, business, computer science, statistics, and all branches of mathematics. A small sample of courses at UW that have linear algebra prerequisites include: AFM 272, ACTSC 291, AMATH 333, CM 271, CO 250, CS 371, ECON 301, PHYS 234, PMATH 330, STAT 331.

Note that although we will mention applications of linear algebra in this course, we will not discuss them in depth. This is because in most cases it would require knowledge of the various applications that we do not expect you to have at this point. Additionally, you will have a lot of time to apply your linear algebra knowledge and skills in future courses.

How to do well in this course!

> *Student: Is there life after death?*
>
> *Teacher: Why do you ask?*
>
> *Student: I'll need the extra time to finish all the homework you gave us.*

1. **Attend all the lectures.**

   Although these course notes have been written to help you learn the course, the lectures are your primary resource to the material covered in the course. The lectures will provide examples and explanations to help you understand the course material. Additionally, the lectures will be interactive. That is, when you do not understand something or need clarification, you can ask. Course notes do not provide this feature (yet). You will likely find it very difficult to get a good grade in the course without attending all the lectures.

2. **Read these course notes.**

   It is generally recommended that students read course notes/text books prior to class. If you are able to teach yourself some of the material before it being taught in class, you will find the lectures twice as effective. The lecturers will then be there for you to clarify what you taught yourself and to help with the areas that you had difficulty with. Trust me, you will enjoy your classes much more if you understand most of what is being taught in the lectures. Additionally, you will likely find it considerably easier to take notes and to listen to the professor at the same time.

3. **Doing/Checking homework assignments.**

   Homework assignments are designed to help you learn the material and prepare for the tests. You should always give a serious effort to solve a problem before seeking help. Spending the time required to solve a hard problem is not only very satisfying, but it will greatly increase your knowledge of the course and your problem solving skills. Naturally, if you are unable to figure out how to solve a problem, it is very important to get help so that you can learn how to do it. Make sure that you always indicate on each assignment where you have received help. **Never just read/copy solutions.** On a test, you are not going to have a solution to read/copy from. You need to make sure that you are able to solve problems without referring to similar problems. Also, the assignments are only worth 5% of your overall mark; the tests, 95%. Struggling on the assignments and learning from your mistakes so that you do much better on the tests is a much better way to go. For this reason, it is highly recommended that you collect your assignments and compare them with the online solutions. Even if you have the correct answer, it is worth checking if there is another way to solve the problem.

4. **Study!**

   This might sound a little obvious, but most students do not do nearly enough of this. Moreover, you need to study properly. If you stay up all night studying drinking Red Bull, the only thing that is going to grow wings and fly away is your knowledge. Learning math, and most other subjects, is like building a house. If you do not have a firm foundation, then it will be very difficult for you to build on top of it. It is extremely important that you know and do not forget the basics.

   For study skills check this out:

   `http://www.adm.uwaterloo.ca/infocs/study/index.html`

5. **Use learning resources.**

   There are many places you can get help outside of the lectures. These include the Tutorial Centre, your lecturer's office hours, the UW-ACE website, and others. They are there to assist you... make use of them!

Proofs

Student 1: What is your favorite part of mathematics?

Student 2: Knot Theory.

Student 1: Me neither.

Do you have difficulty with proof questions? Here are some suggestions to help you.

1. **Know and understand all definitions and theorems.**

   Proofs are essentially puzzles that you have to put together. In class and in these course notes we will give you all of the pieces and it is up to you to figure out how they are to fit together to give the desired picture. Of course, if you are missing or do not understand a required piece, then you are not going to be able to get the correct result.

2. **Learn proof techniques.**

   In class and in the course notes, you will see numerous proofs. You should use these to learn how to select the correct puzzle pieces and to learn the ways of putting the puzzle pieces together. In particular, I always recommend that students ask the following two questions for every step in each proof:

   (a) "Why is the step true?"

   (b) "What was the purpose of the step?"

   Answering the first question should really help your understanding of the definitions and theorems from class (the puzzle pieces). Answering the second question, which is generally more difficult, will help you learn how to think about creating proofs.

3. **Practice!**

   As with everything else, you will get better at proofs with experience. Many proofs in the course notes are left as exercises to give you practice in proving things. It is strongly recommended that you do (try) these proofs before continuing further.

ACKNOWLEDGMENTS

Thanks are expressed to:

Mike La Croix: for formatting, LaTeX consulting, and all of the amazing figures.

MEF: for their generous financial support.

Kate Schreiner and Joseph Horan: for proof-reading, editing, and suggestions.

All of my many colleagues whom have given many useful suggestions on how to teach linear algebra.

Chapter 1

Vectors in Euclidean Space

1.1 Vector Addition and Scalar Multiplication

DEFINITION
$\mathbb{R}^n$

For any positive integer n the set of all elements of the form $(x_1, \ldots, x_n)$ where $x_i \in \mathbb{R}$ for $1 \leq i \leq n$ is called n-dimensional Euclidean space and is denoted $\mathbb{R}^n$. The elements of $\mathbb{R}^n$ are called **points** and are usually denoted $P(x_1, \ldots, x_n)$.

2-dimensional and 3-dimensional Euclidean space was originally studied by the ancient Greek mathematicians. In Euclid's Elements, Euclid defined two and three dimensional space with a set of postulates, definitions, and common notions. However, in modern mathematics, we not only want to make the concepts in Euclidean space more mathematically precise, but we want to be able to easily generalize the concepts to allow us to use the same ideas to solve problems in other areas.

In Linear Algebra, we will view $\mathbb{R}^n$ as a set of **vectors** rather than as a set of points. In particular, we will write an element of $\mathbb{R}^n$ as a column vector and denote it with the usual vector symbol $\vec{x}$ (Note that some textbooks will denote vectors in bold face i.e. $\mathbf{x}$). That is, $\vec{x} \in \mathbb{R}^n$ can be represented as

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R} \text{ for } 1 \leq i \leq n$$

For example, in 3-dimensional Euclidean space, we will denote the origin $(0, 0, 0)$, by the vector $\vec{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

REMARKS

1. Although we are going to view $\mathbb{R}^n$ as a set of vectors, it is important to understand that we really are still referring to n-dimensional Euclidean space. In some cases it will be useful to think of the vector $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ as the point $(x_1, \ldots, x_n)$. In particular, we will sometimes interpret a set of vectors as a set of points to get a geometric object (such as a line or plane).

2. In Linear Algebra all vectors in $\mathbb{R}^n$ should be written as column vectors, however, we will sometimes write these as n-tuples to match the notation that is commonly used in other areas. In particular, when we are dealing with functions of vectors, we will write $f(x_1, \ldots, x_n)$ rather than $f\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right)$.

One advantage in viewing the elements of $\mathbb{R}^n$ as vectors instead of as points is that we can perform operations on vectors. You likely have seen vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ in Physics being used to represent motion or force. From these physical examples we can observe that we add vectors by summing their components and multiply a vector by a scalar by multiplying each entry of the vector by the scalar. We keep this definition for vectors in $\mathbb{R}^n$.

DEFINITION

Vector Addition

Scalar Multiplication

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 $\mathbb{R}^n$ and $c \in \mathbb{R}$. We define

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \qquad c\vec{x} = \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}$$

EXAMPLE 1

Let $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}$ be vectors in $\mathbb{R}^3$. Then,

$$\vec{x} + \vec{y} = \begin{bmatrix} 1 + (-1) \\ 2 + 3 \\ -3 + 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ -3 \end{bmatrix}$$

$$\sqrt{2}\vec{y} = \begin{bmatrix} -\sqrt{2} \\ 3\sqrt{2} \\ 0 \end{bmatrix}$$

$$2\vec{x} - 3\vec{y} = \begin{bmatrix} 2 \\ 4 \\ -6 \end{bmatrix} + \begin{bmatrix} 3 \\ -9 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -5 \\ -6 \end{bmatrix}$$

Since we will often look at the sums of scalar multiples of vectors, we make the following definition.

DEFINITION

Linear Combination

Let $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$. Then the sum

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

where $c_i \in \mathbb{R}$ for $1 \leq i \leq k$ is called a **linear combination** of $\vec{v}_1, \dots, \vec{v}_k$.

Although it is intuitively obvious that for any $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$ that any linear combination

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

is also going to be a vector in $\mathbb{R}^n$, it is instructive to prove this along with several other important properties. We will see throughout Math 136 that these properties are extremely important.

THEOREM 1

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

V1 $\vec{x} + \vec{y} \in \mathbb{R}^n$;
V2 $(\vec{x} + \vec{y}) + \vec{w} = \vec{x} + (\vec{y} + \vec{w})$;
V3 $\vec{x} + \vec{y} = \vec{y} + \vec{x}$;
V4 There exists a vector $\vec{0} \in \mathbb{R}^n$ such that $\vec{x} + \vec{0} = \vec{x}$;
V5 For each $\vec{x} \in \mathbb{R}^n$ there exists a vector $-\vec{x} \in \mathbb{R}^n$ such that $\vec{x} + (-\vec{x}) = \vec{0}$;
V6 $c\vec{x} \in \mathbb{R}^n$;
V7 $c(d\vec{x}) = (cd)\vec{x}$;
V8 $(c + d)\vec{x} = c\vec{x} + d\vec{x}$;
V9 $c(\vec{x} + \vec{y}) = c\vec{x} + c\vec{y}$;
V10 $1\vec{x} = \vec{x}$.

Proof: We will prove V1 and V3 and leave the others as an exercise.

For V1, by definition we have

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

since $x_i + y_i \in \mathbb{R}$.

For V3, we have

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

$\square$

EXERCISE 1 Finish the proof of Theorem 1.

REMARKS

1. Observe that properties V2, V3, V7, V8, V9, and V10 only refer to the operations of addition and scalar multiplication, while the other properties V1, V4, V5, and V6 are about the relationship between the operations and the set $\mathbb{R}^n$. These facts should be clear in the proof of the theorem. Moreover, we see that the zero vector of $\mathbb{R}^n$ is the vector $\vec{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$, and the additive inverse of $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is $-\vec{x} = (-1)\vec{x} = \begin{bmatrix} -x_1 \\ \vdots \\ -x_n \end{bmatrix}$.
2. Properties V1 and V6 show that $\mathbb{R}^n$ is **closed under linear combinations**. That is, if $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$, then $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \in \mathbb{R}^n$ for any $c_1, \ldots, c_k \in \mathbb{R}$. This fact might seem rather obvious in $\mathbb{R}^n$; however, we will soon see that there are sets which are not closed under linear combinations. In Linear Algebra, it will be important and useful to identify which sets have this nice property and, in fact, all the properties V1 - V10.

It can be useful to look at the geometric interpretation of sets of linear combinations of vectors.

EXAMPLE 2 Let $\vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and consider the set S of all scalar multiples of $\vec{v}$. That is,

$$S = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} = c\vec{v} \text{ for some } c \in \mathbb{R}\}$$

What does S represent geometrically?

Solution: Every vector in S has the form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix}, t \in \mathbb{R}$$

Hence, we see this is all vectors in $\mathbb{R}^2$ where $x_1 = x_2$. Alternately, we can view this as the set of all points (x_1, x_1) in $\mathbb{R}^2$, which we recognize as the line $x_2 = x_1$.

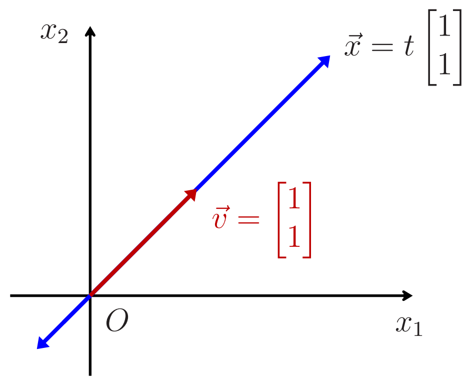

Figure 1.1.1: Geometric representation of S

EXAMPLE 3 Let $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. How would you describe geometrically the set S of all possible linear combinations of $\vec{e}_1$ and $\vec{e}_2$?

Solution: Every vector $\vec{x}$ in S has the form

$$\vec{x} = c_1\vec{e}_1 + c_2\vec{e}_2 = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

for any $c_1, c_2 \in \mathbb{R}$. Therefore, any point in $\mathbb{R}^2$ can be represented as a linear combination of $\vec{e}_1$ and $\vec{e}_2$. Hence, S is the x_1x_2-plane.

Instead of using set notation to indicate a set of vectors, we often instead use the **vector equation** of the set as we did in the example above.

EXAMPLE 4 Let $\vec{m} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. Describe the set with vector equation $\vec{x} = t\vec{m} + \vec{b}$, $t \in \mathbb{R}$ geometrically.

Solution: We have all possible scalar multiples of $\vec{m}$ that are being translated by $\vec{b}$. Hence, this is a line in $\mathbb{R}^3$ that passes through the point $(1, 0, 1)$ that has direction vector $\vec{m}$.

We will often look at the set of all possible linear combinations of a set of vectors. Thus, we make the following definition.

DEFINITION
Span

Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. Then we define the **span** of $\mathcal{B}$ by

$$\operatorname{Span}\mathcal{B} = \{c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k \mid c_1, \dots, c_k \in \mathbb{R}\}$$

We say that $\operatorname{Span}\mathcal{B}$ is **spanned** by $\mathcal{B}$ and that $\mathcal{B}$ is a **spanning set** for $\operatorname{Span}\mathcal{B}$.

DEFINITION
Span

Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. Then we define the **span** of $\mathcal{B}$ by

$$\operatorname{Span}\mathcal{B} = \{c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k \mid c_1, \dots, c_k \in \mathbb{R}\}$$

We say that $\operatorname{Span}\mathcal{B}$ is **spanned** by $\mathcal{B}$ and that $\mathcal{B}$ is a **spanning set** for $\operatorname{Span}\mathcal{B}$.

DEFINITION
Vector Equation

If the set S is spanned by vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$, then a **vector equation** for S is

$$\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k, \quad c_1, \dots, c_k \in \mathbb{R}$$

EXAMPLE 5

Using the definition above, we can write the line in Example 2 as $\operatorname{Span}\left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$.

The set S in Example 3 can be written as $S = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} = \operatorname{Span}\{\vec{e}_1, \vec{e}_2\}$.

EXAMPLE 6

Let $S = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\}$, $T = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}2\\0\\2\end{bmatrix}\right\}$, and $U = \left\{\begin{bmatrix}1\\-1\\2\end{bmatrix}\right\}$. Describe each set geometrically and determine a vector equation for each.

Solution: By definition of span we have

$$S = \left\{ c_1\begin{bmatrix}1\\0\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\\0\end{bmatrix} \mid c_1, c_2 \in \mathbb{R}\right\}.$$

So, we see that S is the set of all vectors of the form $\begin{bmatrix}c_1\\c_2\\0\end{bmatrix}$, and so represents the x_1x_2-plane in $\mathbb{R}^3$. A vector equation for S is $\vec{x} = c_1\begin{bmatrix}1\\0\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\\0\end{bmatrix}$.

For T, we have

$$T = \left\{ c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} \mid c_1, c_2 \in \mathbb{R} \right\}$$

We see that T is the set of all vectors of the form

$$\begin{bmatrix} c_1 + 2c_2 \\ 0 \\ c_1 + 2c_2 \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ c \end{bmatrix}$$

with $c = c_1 + 2c_2$. Hence, T is the line in $\mathbb{R}^3$ with vector equation $\vec{x} = c \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

Observe that U only contains a single vector. Hence, U is just a single point in $\mathbb{R}^3$ and has vector equation $\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$.

We could have also written the vector equation of T as $\vec{x} = c_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$. However, we were able to simplify the vector equation since the second vector was a scalar multiple of the first. In general, when writing a vector equation for a set (or writing the span of a set of vectors), we want to simplify the equation (set) as much as possible. That is, we want to remove any unnecessary vectors.

Let's first consider the simple case of two vectors $\vec{v}_1, \vec{v}_2$ in $\mathbb{R}^3$. Then, we have that $\operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$ is a plane in $\mathbb{R}^3$ if and only if neither $\vec{v}_1$ nor $\vec{v}_2$ is a scalar multiple of the other. What if $\vec{v}_2$ is a scalar multiple of $\vec{v}_1$? Then, we have $\operatorname{Span}\{\vec{v}_1, \vec{v}_2\} = \operatorname{Span}\{\vec{v}_1\}$ and so it either represents a line or just the origin if $\vec{v}_1 = \vec{v}_2 = \vec{0}$.

If we have k vectors in $\mathbb{R}^n$, when do we have

$$\operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_{k-1}\}?$$

For this to be true, the vector $\vec{v}_k$ must be in the set $\operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_{k-1}\}$. Hence, $\vec{v}_k$ must be able to be written as a linear combination of $\vec{v}_1, \dots, \vec{v}_{k-1}$. We now prove the converse.

THEOREM 2

If $\vec{v}_k$ can be written as a linear combination of $\vec{v}_1, \dots, \vec{v}_{k-1}$, then

$$\operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_{k-1}\}$$

Proof: We are assuming that there exist $c_1, \dots, c_{k-1} \in \mathbb{R}$ such that

$$c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} = \vec{v}_k.$$

Let $\vec{x} \in \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$. Then there exist $d_1, \ldots, d_k \in \mathbb{R}$ such that

$$\begin{aligned}\vec{x} &= d_1\vec{v}_1 + \cdots + d_{k-1}\vec{v}_{k-1} + d_k\vec{v}_k \\ &= d_1\vec{v}_1 + \cdots + d_{k-1}\vec{v}_{k-1} + d_k(c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1}) \\ &= (d_1 + d_k c_1)\vec{v}_1 + \cdots + (d_{k-1} + d_k c_{k-1})\vec{v}_{k-1}\end{aligned}$$

Thus, $\vec{x} \in \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$. Hence, $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$. Clearly, we have $\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\} \subseteq \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\}$ and so

$$\operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \operatorname{Span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}$$

as required. $\square$

Of course, by rearranging the vectors in the set we see that any vector that can be written as a linear combination of the other vectors can be removed from the set without changing the set it spans.

EXAMPLE 7

Describe geometrically the following sets and write a simplified vector equation of each.

a) $S = \operatorname{Span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}, \begin{bmatrix}3\\3\end{bmatrix}\right\}$

Solution: By Theorem 2, we have

$$\operatorname{Span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}, \begin{bmatrix}3\\3\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$$

Hence, S is a line in $\mathbb{R}^2$ with vector equation $\vec{x} = c_1\begin{bmatrix}1\\1\end{bmatrix}$.

b) $T = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\0\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$

Solution: Since $\begin{bmatrix}2\\0\end{bmatrix}$ is a scalar multiple of $\begin{bmatrix}1\\0\end{bmatrix}$, Theorem 2 gives

$$\operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\0\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$$

We also have that $\begin{bmatrix}1\\3\end{bmatrix}$ can be written as a linear combination of $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$, hence

$$\operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$$

Hence, T is all of $\mathbb{R}^2$ and has vector equation $\vec{x} = c_1\begin{bmatrix}1\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\end{bmatrix}$.

c) $U = \operatorname{Span}\left\{\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}0\\1\\-1\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}\right\}$

Solution: Since $\begin{bmatrix}0\\0\\0\end{bmatrix}$ is a scalar multiple of any vector we get

$$\text{Span}\left\{\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}0\\1\\-1\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}\right\} = \text{Span}\left\{\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}0\\1\\-1\end{bmatrix}\right\}$$

Hence, U is a plane in $\mathbb{R}^3$ with vector equation $\vec{x} = c_1\begin{bmatrix}1\\1\\2\end{bmatrix} + c_2\begin{bmatrix}0\\1\\-1\end{bmatrix}$.

REMARK

Observe that for T any of the answers

$$\vec{x} = c_1\begin{bmatrix}1\\0\end{bmatrix} + c_2\begin{bmatrix}1\\3\end{bmatrix}$$
$$\vec{x} = c_1\begin{bmatrix}2\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\end{bmatrix}$$
$$\vec{x} = c_1\begin{bmatrix}2\\0\end{bmatrix} + c_2\begin{bmatrix}1\\3\end{bmatrix}$$

would also be correct.

From the examples we see that it is very important to detect when one or more vectors in the set can be made up as linear combinations of some (or all) of the other vectors. We call such sets **linearly dependent sets**. i.e. $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\0\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\}$ is linearly dependent and $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$ is **linearly independent**. In the examples, it was quite easy to see when we had this situation; however, in real world applications we may get hundreds of vectors each having hundreds of entries. So, we need to develop a method for detecting these. To do this, we first need a mathematical definition of linear dependence.

If a set $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly dependent, then it means that one of the vectors can be made as a linear combination of some (or all) of the other vectors. Say, $\vec{v}_i$ can be written as a linear combination of the other vectors, that is, there exist $c_1, \ldots c_k \in \mathbb{R}$ such that

$$-c_i v_i = c_1 v_1 + \cdots + c_{i-1}v_{i-1} + c_{i+1}v_{i+1} + \cdots + c_k v_k$$

where $c_i \neq 0$. We can rewrite this as

$$\vec{0} = c_1 v_1 + \cdots + c_{i-1}v_{i-1} + c_i v_i + c_{i+1}v_{i+1} + \cdots + c_k v_k$$

Hence, we get our mathematical definition.

DEFINITION

Linearly Dependent

Linearly Independent

A set of vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ in $\mathbb{R}^n$ is said to be **linearly dependent** if there exist coefficients $c_1, \dots, c_k$ not all zero such that

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$$

Thus, a set of vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ is said to be **linearly independent** if the only solution to

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$$

is $c_1 = c_2 = \cdots = c_k = 0$ (called the **trivial solution**).

THEOREM 3

If a set of vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ contains the zero vector then it is linearly dependent.

Proof: Without loss of generality assume $\vec{v}_k = \vec{0}$. Then we have

$$0\vec{v}_1 + 0\vec{v}_2 + \cdots + 0\vec{v}_{k-1} + 1\vec{v}_k = \vec{0}$$

Hence, the equation $\vec{0} = c_1 v_1 + \cdots + c_k v_k$ has a solution with one coefficient, namely c_k, that is non-zero. So, by definition, the set is linearly dependent. $\square$

EXAMPLE 8

Determine whether $\left\{ \begin{bmatrix} 7 \\ -14 \\ 6 \end{bmatrix}, \begin{bmatrix} -10 \\ 15 \\ 15/14 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \right\}$ is linearly dependent or linearly independent.

Solution: We consider

$$c_1 \begin{bmatrix} 7 \\ -14 \\ 6 \end{bmatrix} + c_2 \begin{bmatrix} -10 \\ 15 \\ 15/14 \end{bmatrix} + c_3 \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Using operations on vectors we get

$$\begin{bmatrix} 7c_1 - 10c_2 - c_3 \\ -14c_1 + 15c_2 \\ 6c_1 + \frac{15}{14}c_2 + 3c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Thus, we get 3 equations in 3 unknowns

$$\begin{aligned} 7c_1 - 10c_2 - c_3 &= 0 \\ -14c_1 + 15c_2 \qquad &= 0 \\ 6c_1 + \frac{15}{14}c_2 + 3c_3 &= 0 \end{aligned}$$

Solving we find that there are in fact infinitely many possible solutions. One is $c_1 = \frac{3}{7}$, $c_2 = \frac{2}{5}$ and $c_3 = -1$. Hence, the set is linearly dependent.

REMARK

Observe that for determining whether a set $\{\vec{v}_1, \ldots, \vec{v}_k\}$ in $\mathbb{R}^n$ is linearly dependent or linearly independent requires determining solutions of the vector equation $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \vec{0}$. However, this equation actually represents n equations (one for each entry of the vectors) in k unknowns $c_1, \ldots, c_k$. In the next chapter, we will look at how to efficiently solve such systems of equations.

What we have derived above is that the simplest spanning set for a given set is one that is linearly independent. Hence, we make the following definition.

DEFINITION
Basis

If a subset S of $\mathbb{R}^n$ can be written as a span of vectors $\vec{v}_1, \ldots, \vec{v}_k$ where $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent, then we say that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a **basis** for S.

This definition and its generalization will be extremely important throughout the remainder of the book. At this point, however, we just define the following very important bases.

DEFINITION
Standard Basis

In $\mathbb{R}^n$, let $\vec{e}_i$ represent the vector whose i-th component is 1 and all other components are 0. The set $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$.

EXAMPLE 9

The standard basis for $\mathbb{R}^2$ is $\{\vec{e}_1, \vec{e}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$.

EXERCISE 2

Write the standard basis for $\mathbb{R}^4$.

Surfaces in Higher Dimensions

We can extend our geometrical concepts of lines and planes to $\mathbb{R}^n$ for $n > 3$.

DEFINITION
Line in $\mathbb{R}^n$

Let $\vec{v}, \vec{b} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$. Then we call the set with vector equation $\vec{x} = c_1\vec{v} + \vec{b}$, $c_1 \in \mathbb{R}$ a **line** in $\mathbb{R}^n$ which passes through $\vec{b}$.

DEFINITION
Plane in $\mathbb{R}^n$

Let $\vec{v}_1, \vec{v}_2, \vec{b} \in \mathbb{R}^n$ with $\{\vec{v}_1, \vec{v}_2\}$ being a linearly independent set. Then we call the set with vector equation $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \vec{b}$, $c_1, c_2 \in \mathbb{R}$ a **plane** in $\mathbb{R}^n$ which passes through $\vec{b}$.

DEFINITION
k-plane in $\mathbb{R}^n$

Let $\vec{v}_1, \dots, \vec{v}_k, \vec{b} \in \mathbb{R}^n$ with $\{\vec{v}_1, \dots, \vec{v}_k\}$ being a linearly independent set. Then we call the set with vector equation $\vec{x} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k + \vec{b}$, $c_1, \dots, c_k \in \mathbb{R}$ a ***k*-plane** in $\mathbb{R}^n$ which passes through $\vec{b}$.

DEFINITION
Hyperplane in $\mathbb{R}^n$

Let $\vec{v}_1, \dots, \vec{v}_{n-1}, \vec{b} \in \mathbb{R}^n$ with $\{\vec{v}_1, \dots, \vec{v}_{n-1}\}$ being linearly independent. Then the set with vector equation $\vec{x} = c_1\vec{v}_1 + \cdots + c_{n-1}\vec{v}_{n-1} + \vec{b}$, $c_i \in \mathbb{R}$ is called a **hyperplane** in $\mathbb{R}^n$ which passes through $\vec{b}$.

REMARK

Geometrically we see that a line in $\mathbb{R}^n$ has dimension 1, a k-plane in $\mathbb{R}^n$ has dimension k, and a hyperplane in $\mathbb{R}^n$ has dimension $n-1$. We will make a precise mathematical definition of dimension in Chapter 4.

EXAMPLE 10

The set Span $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix} \right\}$ defines a hyperplane in $\mathbb{R}^4$ since $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix} \right\}$ is linearly independent.

EXERCISE 3

Give a vector equation for a hyperplane in $\mathbb{R}^5$.

1.2 Subspaces

As we mentioned when we first looked at the ten properties of vector addition and scalar multiplication, Theorem 1, properties V1 and V6 seem rather obvious. However, it is easy to see that not all subsets of $\mathbb{R}^n$ have this property. For example, the set $S = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ clearly does satisfy property V1 and V6 since

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \notin S$$
$$c\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} c \\ c \end{bmatrix} \notin S \text{ if } c \neq 1$$

A set which satisfies property V1 is said to be **closed under addition**. A set which satisfies property V6 is said to be **closed under scalar multiplication**.

You might be wondering why we should be so interested in whether a set has properties V1 and V6 or not. One reason is that a non-empty subset of $\mathbb{R}^n$ which does not have these properties is not closed under linear combinations. Another reason is that a non-empty subset of $\mathbb{R}^n$ which does have these properties will in fact satisfy all the same ten properties as $\mathbb{R}^n$. That is, it is a "space" which is a subset of the larger space $\mathbb{R}^n$.

DEFINITION

Subspace of $\mathbb{R}^n$

A non-empty subset $\mathbb{S}$ of $\mathbb{R}^n$ is called a **subspace of** $\mathbb{R}^n$ if it satisfies all ten properties V1 - V10 of Theorem 1.

It, of course, would be tedious to have to check all ten properties each time we want to prove a non-empty subset $\mathbb{S}$ of $\mathbb{R}^n$ is a subspace. The good news is that this is not necessary. Observe that properties V2, V3, V7, V8, V9, V10 are only about the operations of addition and scalar multiplication. Thus, we already know that these properties hold. Also, if V1 and V6 hold, then for any $\vec{v} \in \mathbb{S}$ we have $\vec{0} = 0\vec{v} \in \mathbb{S}$ and $(-1)\vec{v} = (-\vec{v}) \in \mathbb{S}$, so V4 and V5 hold. Thus, to check if a non-empty subset of $\mathbb{R}^n$ is a subspace, we can use the following result.

THEOREM 1

(Subspace Test)

Let $\mathbb{S}$ be a non-empty subset of $\mathbb{R}^n$. If $\vec{x}+\vec{y} \in \mathbb{S}$ and $c\vec{x} \in \mathbb{S}$ for all $\vec{x}, \vec{y} \in \mathbb{S}$ and $c \in \mathbb{R}$, then $\mathbb{S}$ is a subspace of $\mathbb{R}^n$.

REMARK

If $\mathbb{S}$ is a non-empty and is closed under scalar multiplication, then our work above shows us that $\mathbb{S}$ must contain $\vec{0}$. Thus, the standard way of checking if $\mathbb{S}$ is non-empty is to determine if $\vec{0}$ satisfies the conditions of the set. If $\vec{0}$ is not in $\mathbb{S}$, then we automatically know that it is not a subspace of $\mathbb{R}^n$ as it would not satisfy V4 or V6.

EXAMPLE 1

Determine, with proof, which of the following are subspaces of $\mathbb{R}^3$. Describe each subspace geometrically and write a basis for the subspace.

a) $\mathbb{S}_1 = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mid x_1 + x_2 = 0, x_1 - x_3 = 0 \right\}$.

Solution: By definition $\mathbb{S}_1$ is a subset of $\mathbb{R}^3$, and $\vec{0}$ satisfies the conditions of the set $(0+0=0$ and $0-0=0)$ so $\vec{0} \in \mathbb{S}_1$. Thus, $\mathbb{S}_1$ is non-empty subset of $\mathbb{R}^3$, so we can apply the Subspace Test.

To show that $\mathbb{S}_1$ is closed under addition we pick two vectors $\vec{x}, \vec{y} \in \mathbb{S}_1$ and show that $\vec{x}+\vec{y}$ satisfies the conditions of $\mathbb{S}_1$. If $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ are in $\mathbb{S}_1$, then we have $x_1+x_2 = 0$, $x_1 - x_3 = 0$, $y_1 + y_2 = 0$, and $y_1 - y_3 = 0$. By definition of addition, we get

$$\vec{x}+\vec{y} = \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ x_3+y_3 \end{bmatrix}$$

Then we see that

$$(x_1+y_1)+(x_2+y_2) = x_1+x_2+y_1+y_2 = 0+0 = 0$$

and

$$(x_1+y_1)-(x_3+y_3) = x_1-x_3+y_1-y_3 = 0+0 = 0$$

Hence, $\vec{x}+\vec{y}$ satisfies the conditions of $\mathbb{S}_1$, so $\mathbb{S}_1$ is closed under addition.

Similarly, for any $c \in \mathbb{R}$ we have

$$c\vec{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ cx_3 \end{bmatrix}$$

and

$$cx_1 + cx_2 = c(x_1+x_2) = c0 = 0, \quad cx_1 - cx_3 = c(x_1 - x_3) = c0 = 0$$

Hence, $\mathbb{S}_1$ is also closed under scalar multiplication.

Therefore, by the subspace test $\mathbb{S}_1$ is a subspace of $\mathbb{R}^3$.

To describe $\mathbb{S}_1$ geometrically, we see that since $x_1+x_2 = 0$ and $x_1 - x_3 = 0$ we have $x_2 = -x_1$ and $x_3 = x_1$. Thus, a general vector in the set can be represented as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 \\ x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

Thus, $\mathbb{S}_1$ is a line in $\mathbb{R}^3$ and $\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{S}_1$.

b) $\mathbb{S}_2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mid x_1 + x_2 = 1 \right\}$.

Solution: By definition $\mathbb{S}_2$ is a subset of $\mathbb{R}^3$, but we see that $\vec{0}$ does not satisfy the condition of $\mathbb{S}_2$ $(0+0 \neq 1)$, hence $\mathbb{S}_2$ is not a subspace of $\mathbb{R}^3$.

c) $\mathbb{S}_3 = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mid x_1 - 2x_2 = 0 \right\}$.

Solution: By definition $\mathbb{S}_3$ is a subset of $\mathbb{R}^3$, and $\vec{0}$ satisfies the conditions of the set $(0 - 2(0) = 0)$ so $\vec{0} \in \mathbb{S}_3$. Thus, $\mathbb{S}_3$ is non-empty subset of $\mathbb{R}^3$, so we can apply the Subspace Test.

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 vectors in $\mathbb{S}_3$, then we have $x_1 - 2x_2 = 0$, and $y_1 - 2y_2 = 0$. Then

$$\vec{x} + \vec{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}$$

and

$$(x_1 + y_1) - 2(x_2 + y_2) = x_1 - 2x_2 + y_1 - 2y_2 = 0 + 0 = 0$$

Hence, $\mathbb{S}_3$ is closed under addition.

Similarly, for any $c \in \mathbb{R}$ we have

$$c\vec{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ cx_3 \end{bmatrix}$$

and

$$cx_1 - 2(cx_2) = c(x_1 - 2x_2) = c0 = 0$$

Hence, $\mathbb{S}_3$ is also closed under scalar multiplication.

Therefore, $\mathbb{S}_3$ is a subspace of $\mathbb{R}^3$ by the Subspace Test.

Since $x_1 - 2x_2 = 0$ we have $x_1 = 2x_2$. So, a general vector in $\mathbb{S}_3$ can be represented as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Thus, $\mathbb{S}_3$ is a plane in $\mathbb{R}^3$ since $\mathcal{B}_2 = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{S}_3$.

EXERCISE 1

Let $\vec{v}, \vec{b} \in \mathbb{R}^n$. Show that the line $\mathbb{S} = \{t\vec{v} + \vec{b} \mid t \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^n$ if and only if $\vec{b}$ is a scalar multiple of $\vec{v}$.

THEOREM 2

Let $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$. Then $\mathbb{S} = \operatorname{Span}\{\vec{v}_1, \dots, \vec{v}_k\}$ is a subspace of $\mathbb{R}^n$.

Proof: The proof is left as an exercise.

1.3 Dot Product

In high school you saw the dot product in $\mathbb{R}^2$ and $\mathbb{R}^3$. Recall that

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1y_1 + x_2y_2$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1y_1 + x_2y_2 + x_3y_3$$

You saw that two useful aspects of the dot product were in calculating the length of a vector and in determining whether two vectors are orthogonal. In particular, the length of a vector $\vec{v}$ is $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$, and two vectors $\vec{x}$ and $\vec{y}$ are orthogonal if and only if $\vec{x} \cdot \vec{y} = 0$. In relation to the latter, the dot product also determines the angle between two vectors in $\mathbb{R}^2$.

THEOREM 1

Let $\vec{x}, \vec{y} \in \mathbb{R}^2$ and let θ be the angle between $\vec{x}$ and $\vec{y}$. Then

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

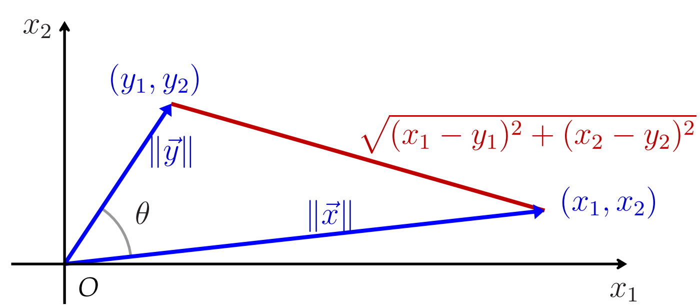

Figure 1.3.1: A figure illustrating the relationship between the dot product and cosines.

Since the dot product is so useful in $\mathbb{R}^2$ and $\mathbb{R}^3$, it makes sense to extend it to $\mathbb{R}^n$.

DEFINITION
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 $\mathbb{R}^n$. Then the **dot product** of $\vec{x}$ and $\vec{y}$ is

$$\vec{x} \cdot \vec{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i$$

REMARK

The dot product is also called the **standard inner product** or the **scalar product** of $\mathbb{R}^n$.

EXAMPLE 1

Let $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, and $\vec{z} = \begin{bmatrix} 3 \\ 1 \\ -2 \\ -2 \end{bmatrix}$. Calculate $\vec{x} \cdot \vec{y}$, $\vec{y} \cdot \vec{z}$, $\vec{z} \cdot \vec{z}$, and $(\vec{x} \cdot \vec{z})\vec{z}$.

Solution: We have

$$\begin{aligned} \vec{x} \cdot \vec{y} &= 1(0) + 2(0) + (-1)(1) + 1(1) = 0 \\ \vec{y} \cdot \vec{z} &= 0(3) + 0(1) + 1(-2) + 1(-2) = -4 \\ \vec{z} \cdot \vec{z} &= 3(3) + 1(1) + (-2)(-2) + (-2)(-2) = 18 \\ (\vec{x} \cdot \vec{z})\vec{z} &= \big(1(3) + 2(1) + (-1)(-2) + 1(-2)\big)\vec{z} = 5 \begin{bmatrix} 3 \\ 1 \\ -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 15 \\ 5 \\ -10 \\ -10 \end{bmatrix} \end{aligned}$$

THEOREM 2

Let $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ and let $s, t \in \mathbb{R}$. Then

(1) $\vec{x} \cdot \vec{x} \geq 0$ and $\vec{x} \cdot \vec{x} = 0$ if and only if $\vec{x} = \vec{0}$.
(2) $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$
(3) $\vec{x} \cdot (s\vec{y} + t\vec{z}) = s(\vec{x} \cdot \vec{y}) + t(\vec{x} \cdot \vec{z})$

Proof: Let $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, and $\vec{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$. Then for (1)

$$\vec{x} \cdot \vec{x} = x_1^2 + \cdots + x_n^2 \geq 0$$

and $\vec{x} \cdot \vec{x} = 0$ if and only if $x_i = 0$ for $1 \leq i \leq n$.

For (2) we have

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

For (3) We have

$$\begin{aligned}\vec{x} \cdot (s\vec{y} + t\vec{z}) &= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} sy_1 + tz_1 \\ \vdots \\ sy_n + tz_n \end{bmatrix} \\ &= x_1(sy_1 + tz_1) + \cdots + x_n(sy_n + tz_n) \\ &= sx_1y_1 + tx_1z_1 + \cdots + sx_ny_n + tx_nz_n \\ &= s(x_1y_1 + \cdots + x_ny_n) + t(x_1z_1 + \cdots + x_nz_n) \\ &= s(\vec{x} \cdot \vec{y}) + t(\vec{x} \cdot \vec{z})\end{aligned}$$

□

Observe that property (1) allows us to define the length of a vector in $\mathbb{R}^n$ to match our formula in $\mathbb{R}^2$ and $\mathbb{R}^3$.

DEFINITION
Length

Let $\vec{x} \in \mathbb{R}^n$. The **length** or **norm** of $\vec{x}$ is defined to be

$$\|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{x}}$$

EXAMPLE 2

Find the length of $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ 1 \\ 1 \end{bmatrix}$.

Solution: We have

$$\begin{aligned}\|\vec{x}\| &= \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{1^2 + 2^2 + (-1)^2 + 0^2} = \sqrt{6} \\ \|\vec{y}\| &= \sqrt{\vec{y} \cdot \vec{y}} = \sqrt{1^2 + (-2)^2 + 3^2 + 1^2 + 1^2} = \sqrt{16} = 4\end{aligned}$$

DEFINITION
Unit Vector

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

THEOREM 3

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

(1) $\|\vec{x}\| \geq 0$ and $\|\vec{x}\| = 0$ if and only if $\vec{x} = \vec{0}$.
(2) $\|c\vec{x}\| = |c|\|\vec{x}\|$
(3) $|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$ (Cauchy-Schwarz-Buniakowski Inequality)
(4) $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (Triangle Inequality)

Proof: We prove (3) and leave the rest as exercises.

If $\vec{x} = \vec{0}$ the result is obvious. For $\vec{x} \neq \vec{0}$, we use (1) and properties of dot products to get

$$0 \leq (t\vec{x} + \vec{y}) \cdot (t\vec{x} + \vec{y}) = (\vec{x} \cdot \vec{x})t^2 + 2(\vec{x} \cdot \vec{y})t + (\vec{y} \cdot \vec{y})$$

for any $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $t \in \mathbb{R}$. This is a polynomial in t where the coefficient $\vec{x} \cdot \vec{x}$ of t^2 is positive. Thus, for the polynomial to be non-negative it must have non-positive discriminant. Hence,

$$\begin{aligned}(2\vec{x} \cdot \vec{y})^2 - 4(\vec{x} \cdot \vec{x})(\vec{y} \cdot \vec{y}) &\geq 0 \\ 4(\vec{x} \cdot \vec{y})^2 - 4\|\vec{x}\|^2\|\|\vec{y}\|^2 &\geq 0 \\ (\vec{x} \cdot \vec{y})^2 &\geq \|\vec{x}\|^2\|\vec{y}\|^2\end{aligned}$$

Taking square roots of both sides gives the desired result. $\square$

We now also extend the concept of angles, and in particular orthogonality, to vectors in $\mathbb{R}^n$.

DEFINITION
Angle in $\mathbb{R}^n$

Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then we define the **angle** between $\vec{x}$ and $\vec{y}$ to be the angle θ such that

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

REMARK

Observe that property (3) of Theorem 3 guarantees that such an angle θ exists for each pair $\vec{x}$, $\vec{y}$.

DEFINITION
Orthogonal

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

EXAMPLE 3

The vectors $\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 2 \\ 3 \\ -4 \\ 1 \end{bmatrix}$ are orthogonal because

$$\vec{x} \cdot \vec{y} = 1(2) + 0(3) + 1(-4) + 2(1) = 0$$

The vectors $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 5 \end{bmatrix}$ are not orthogonal becuase

$$\vec{u} \cdot \vec{v} = 1(1) + 2(1) + 1(2) + (-1)(1) + 5(1) = 9$$

The zero vector $\vec{0}$ in $\mathbb{R}^n$ is orthogonal to every vector $\vec{x} \in \mathbb{R}^n$ since $\vec{x} \cdot \vec{0} = 0$.

EXAMPLE 4

The standard basis vectors for $\mathbb{R}^3$ form an **orthogonal set** since every pair of vectors in the set is orthogonal. That is

$$\begin{bmatrix}1\\0\\0\end{bmatrix} \cdot \begin{bmatrix}0\\1\\0\end{bmatrix} = 0$$
$$\begin{bmatrix}1\\0\\0\end{bmatrix} \cdot \begin{bmatrix}0\\0\\1\end{bmatrix} = 0$$
$$\begin{bmatrix}0\\0\\1\end{bmatrix} \cdot \begin{bmatrix}0\\1\\0\end{bmatrix} = 0$$

REMARK

In fact, one reason the standard basis vectors for $\mathbb{R}^n$ are so easy to work with is because that they form an orthogonal set of unit vectors. This will be examined in Math 235.

Scalar Equation of a Plane

We can use the dot product to write the equation of a plane in $\mathbb{R}^3$ in a useful form. Let $\vec{n} = \begin{bmatrix}n_1\\n_2\\n_3\end{bmatrix}$ be a vector in $\mathbb{R}^3$ that is orthogonal to every vector in the plane (called the **normal vector**) and let $A(a_1, a_2, a_3)$ be any point on the plane. Then for any other point $X(x_1, x_2, x_3)$ in the plane we can form the **directed line segment** between A and X. We define

$$\overrightarrow{AX} = \begin{bmatrix}x_1 - a_1\\x_2 - a_2\\x_3 - a_3\end{bmatrix}$$

By definition, this is a vector which lies in the plane and hence is orthogonal to $\vec{n}$. Thus we have $\vec{n} \cdot \overrightarrow{AX} = 0$. Simplifying this we get

$$\begin{bmatrix}n_1\\n_2\\n_3\end{bmatrix} \cdot \begin{bmatrix}x_1 - a_1\\x_2 - a_2\\x_3 - a_3\end{bmatrix} = 0$$
$$n_1(x_1 - a_1) + n_2(x_2 - a_2) + n_3(x_3 - a_3) = 0$$
$$n_1x_1 + n_2x_2 + n_3x_3 = n_1a_1 + n_2a_2 + n_3a_3$$

Observe that this equation is valid for every point $X(x_1, x_2, x_3)$ in the plane, and hence is an equation of the plane. It is called the **scalar equation** of the plane in $\mathbb{R}^3$.

EXAMPLE 5

Find the scalar equation of the plane which passes through $A(1, 2, 3)$ with normal vector $\vec{n} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$.

Solution: From our work above, we have that the scalar equation is

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

EXAMPLE 6

Find the scalar equation of the plane which passes through $A(-1, 0, 2)$ that is parallel to the plane $x_1 - 2x_2 + 3x_3 = -2$.

Solution: For two planes to be parallel, their normal vectors must also be parallel. That is, a normal vector for the required plane is also $\vec{n} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$. Thus, the scalar equation of the required plane is

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

We can extend this to $\mathbb{R}^m$. Let $\vec{n} = \begin{bmatrix} n_1 \\ \vdots \\ n_m \end{bmatrix}$. The set of all vectors which are orthogonal to $\vec{n}$ will form a hyperplane of $\mathbb{R}^m$. Thus, the standard form for the scalar equation of a hyperplane in $\mathbb{R}^m$ is

$$n_1x_1 + \cdots + n_mx_m = d$$

where d is a constant.

Notice that our derivation of the scalar equation required us to know the normal vector of the plane in advance. However, as we saw in the previous section, typically in Linear Algebra we will be given the vector equation of a plane. Therefore, we need to determine how to find a normal vector of a plane given its vector equation.

Let P be the plane with vector equation

$$\vec{x} = c_1\vec{v} + c_2\vec{w} + \vec{b} = c_1 \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} + c_2 \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$c_1, c_2 \in \mathbb{R}$. We want to find a vector $\vec{n}$ that is orthogonal to every vector which lies in the plane. We first find directed line segments which lie in the plane that originate from the same point. Taking $c_1 = 0 = c_2$ we get that $B(b_1, b_2, b_3)$ lies in the plane, taking $c_1 = 1$, $c_2 = 0$ we get that $V(v_1 + b_1, v_2 + b_2, v_3 + b_3)$ lies in the plane, and taking $c_1 = 0$, $c_2 = 1$, we get that $W(w_1+b_1, w_2+b_2, w_3+b_3)$ lies in the plane. Hence, the directed line segments

$$\overrightarrow{BV} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \vec{v}$$

$$\overrightarrow{BW} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \vec{w}$$

are vectors in the plane which originate from the same point $\vec{b}$. We now want to find a vector $\vec{n}$ such that

$$\vec{n} \cdot \vec{v} = 0 = \vec{n} \cdot \vec{w}$$

This gives the system of two equations in three unknowns

$$v_1 n_1 + v_2 n_2 + v_3 n_3 = 0$$
$$w_1 n_1 + w_2 n_2 + w_3 n_3 = 0$$

Solving the system we find that one solution is

$$\vec{n} = \begin{bmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{bmatrix}$$

DEFINITION
Cross Product

Let $\vec{v}, \vec{w} \in \mathbb{R}^3$. Then, the **cross product** of $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$ is

$$\vec{v} \times \vec{w} = \begin{bmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{bmatrix}$$

EXAMPLE 7

We have

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix} = \begin{bmatrix} 2(4) - 3(5) \\ 3(6) - 1(4) \\ 1(5) - 2(6) \end{bmatrix} = \begin{bmatrix} -7 \\ 14 \\ -7 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1(0) - 0(0) \\ 0(1) - 0(0) \\ 0(0) - 1(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

By definition, the cross product of $\vec{v}$ and $\vec{w}$ gives a vector $\vec{n}$ that is orthogonal to both $\vec{v}$ and $\vec{w}$. Additionally, the cross product has the following properties.

THEOREM 4

For any vectors $\vec{v}, \vec{w}, \vec{x} \in \mathbb{R}^3$ and $c \in \mathbb{R}$ we have

(1) If $\vec{n} = \vec{v} \times \vec{w}$, then for any $\vec{y} \in \operatorname{Span}\{\vec{v}, \vec{w}\}$ we have $\vec{y} \cdot \vec{n} = 0$.
(2) $\vec{v} \times \vec{w} = -\vec{w} \times \vec{v}$
(3) $\vec{v} \times \vec{v} = \vec{0}$
(4) If $\vec{v} \times \vec{w} = \vec{0}$ then either one of $\vec{v}$ or $\vec{w}$ is the zero vector or $\vec{w}$ is a scalar multiple of $\vec{v}$.
(5) $\vec{v} \times (\vec{w} + \vec{x}) = \vec{v} \times \vec{w} + \vec{v} \times \vec{x}$
(6) $(k\vec{v}) \times (\vec{w}) = k(\vec{v} \times \vec{w})$.
(7) $\|\vec{v} \times \vec{w}\| = \big|\|\vec{v}\|\|\vec{w}\| \sin\theta\big|$ where θ is the angle between $\vec{v}$ and $\vec{w}$.

Proof: The proof is left as an exercise.

EXAMPLE 8

Find a scalar equation for the plane with vector equation $\vec{x} = s\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} + t\begin{bmatrix} -1 \\ -1 \\ 5 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $s, t \in \mathbb{R}$.

Solution: From our work above, we see that a normal vector to the plane is

$$\vec{n} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} \times \begin{bmatrix} -1 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ -14 \\ -2 \end{bmatrix}$$

Thus, a scalar equation of the plane is

$$4x_1 + (-14)x_2 + (-2)x_3 = 4(1) + (-14)(2) + (-2)(3) = -30$$

1.4 Projections

So far we have seen how to combine vectors together (via linear combinations) to create other vectors. We now look at this procedure in reverse. We want to take a vector and write it as a sum of two other vectors. In elementary physics, this is used to split a force into its components along certain directions (usually horizontally and vertical).

In particular, given vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$ we want to write $\vec{u}$ as a sum of a scalar multiple of $\vec{v}$ and another vector $\vec{w}$ which is orthogonal to $\vec{v}$ as in Figure 1.4.1. To do this, we first need to find how much of $\vec{u}$ is in the direction of $\vec{v}$.

Consider $\vec{u} = c\vec{v} + \vec{w}$. Then

$$\vec{u} \cdot \vec{v} = (c\vec{v} + \vec{w}) \cdot \vec{v} = c(\vec{v} \cdot \vec{v}) + \vec{w} \cdot \vec{v} = c\|\vec{v}\|^2 + 0$$

Hence, $c = \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2}$.

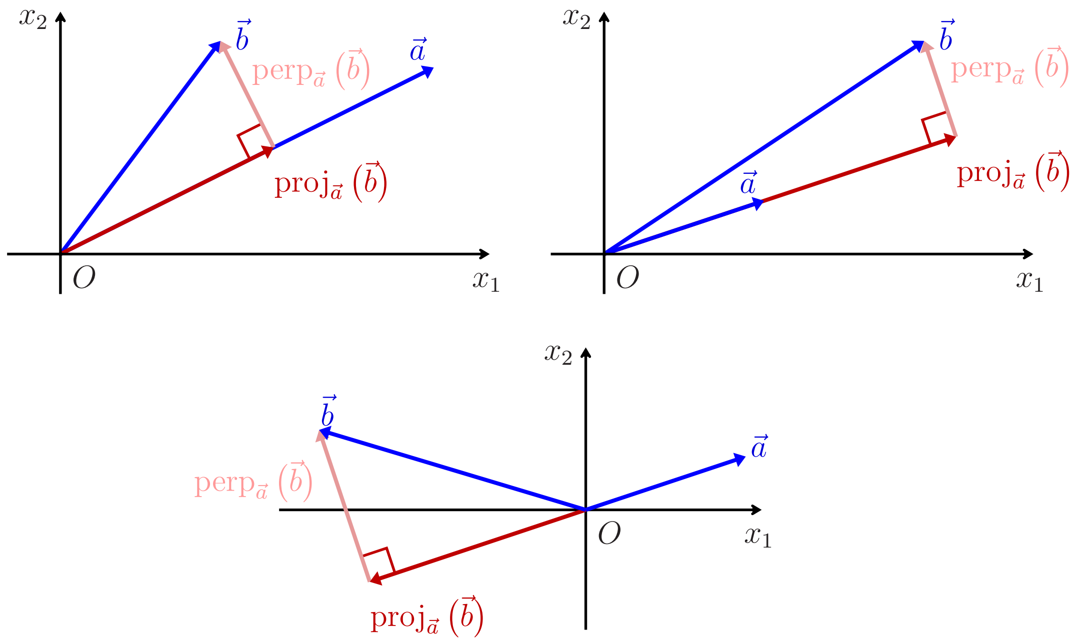

Figure 1.4.1: Examples of projections

DEFINITION
Projection in $\mathbb{R}^n$

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

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

Now that we have the projection of $\vec{u}$ onto $\vec{v}$, it is easy to find the vector $\vec{w}$ that is orthogonal to $\vec{v}$ such that $\vec{u} = \vec{v} + \vec{w}$.

DEFINITION
Perpendicular in $\mathbb{R}^n$

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

$$\text{perp}_{\vec{v}}\, \vec{u} = \vec{u} - \text{proj}_{\vec{v}}\, \vec{u}$$

EXERCISE 1

Show that $\text{proj}_{\vec{v}}\, \vec{u} \cdot \text{perp}_{\vec{v}}\, \vec{u} = 0$.

EXAMPLE 1

Let $\vec{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Find $\text{proj}_{\vec{v}}\, \vec{u}$ and $\text{perp}_{\vec{v}}\, \vec{u}$.

Solution: We have

$$\text{proj}_{\vec{v}}\,\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{18}{25}\begin{bmatrix}3\\4\end{bmatrix}$$
$$\text{perp}_{\vec{v}}(\vec{u}) = \vec{u} - \text{proj}_{\vec{v}}\,\vec{u} = \begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}54/25\\72/25\end{bmatrix} = \begin{bmatrix}-4/25\\3/25\end{bmatrix}$$

EXAMPLE 2 Let $\vec{v} = \begin{bmatrix}1\\2\\2\end{bmatrix}$ and $\vec{u} = \begin{bmatrix}1\\0\\1\end{bmatrix}$. Find $\text{proj}_{\vec{v}}\,\vec{u}$ and $\text{perp}_{\vec{v}}\,\vec{u}$.

Solution: We have

$$\text{proj}_{\vec{v}}\,\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\|\vec{v}\|^2}\vec{v} = \frac{3}{9}\begin{bmatrix}1\\2\\2\end{bmatrix} = \begin{bmatrix}1/3\\2/3\\2/3\end{bmatrix}$$
$$\text{perp}_{\vec{v}}(\vec{u}) = \vec{u} - \text{proj}_{\vec{v}}\,\vec{u} = \begin{bmatrix}1\\0\\1\end{bmatrix} - \begin{bmatrix}1/3\\2/3\\2/3\end{bmatrix} = \begin{bmatrix}2/3\\-2/3\\1/3\end{bmatrix}$$

How do we find the projection of a vector onto a plane? The problem is that the plane is made up of two linearly independent vectors, so we don't really know onto which vector in the plane we are projecting. From Figure 1.4.2 we see that we can instead use the normal vector of the plane. In particular, the projection of the vector onto the plane is the perpendicular of the projection of the vector onto the normal vector of the plane.

EXAMPLE 3 Find the projection of $\vec{u} = \begin{bmatrix}2\\3\\1\end{bmatrix}$ onto the plane $3x_1 - x_2 + 4x_3 = 0$.

Solution: We see that the normal vector is $\vec{n} = \begin{bmatrix}3\\-1\\4\end{bmatrix}$. From our work above the projection of $\vec{u}$ onto the plane is the perpendicular of the projection of $\vec{u}$ onto $\vec{n}$. That is

$$\text{perp}_{\vec{n}}\,\vec{u} = \vec{u} - \frac{\vec{u}\cdot\vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix}2\\3\\1\end{bmatrix} - \frac{7}{26}\begin{bmatrix}3\\-1\\4\end{bmatrix} = \begin{bmatrix}31/26\\85/26\\-1/13\end{bmatrix}$$

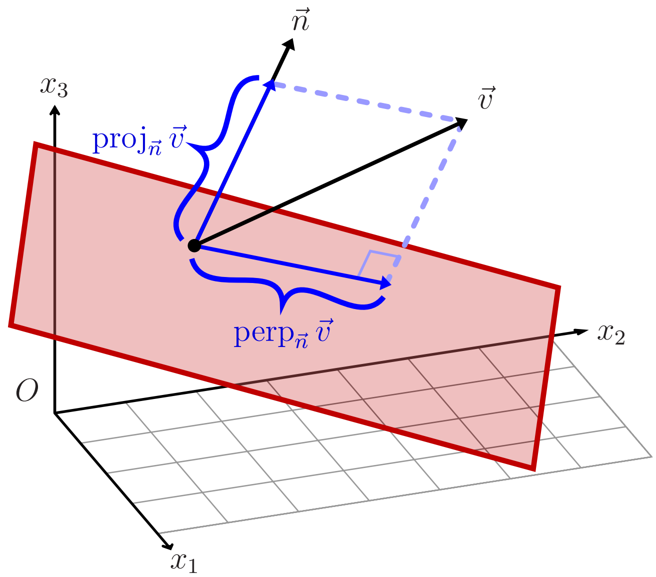

Figure 1.4.2: projection of a vector onto a plane

EXAMPLE 4

Find the projection of $\vec{u} = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}$ onto the plane with vector equation $\vec{x} = s\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + t\begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix}$, $s, t \in \mathbb{R}$.

Solution: To find the projection, we first need to find a normal vector of the plane. We have

$$\vec{n} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \times \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 6 \end{bmatrix}$$

Thus, the projection of $\vec{u}$ onto the plane is

$$\text{perp}_{\vec{n}}\,\vec{u} = \vec{u} - \frac{\vec{u}\cdot\vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} - \frac{18}{44}\begin{bmatrix} -2 \\ -2 \\ 6 \end{bmatrix} = \begin{bmatrix} -2/11 \\ 20/11 \\ 6/11 \end{bmatrix}$$

EXAMPLE 5

Find the projection of $\vec{u} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$ onto the plane $P = \text{Span}\left\{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right\}$.

Solution: The vector equation of the plane is $\vec{x} = s\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $s, t \in \mathbb{R}$. Thus, a normal vector of the plane is

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

Thus, the projection of $\vec{u}$ onto P is

$$\operatorname{perp}_{\vec{n}} \vec{u} = \vec{u} - \frac{\vec{u} \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} - \frac{15}{9} \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$

REMARK

In the last two examples we could have used any scalar multiple of $\vec{n}$ as a normal vector.

Chapter 2

Systems of Linear Equations

We have already seen in the last chapter many cases where we needed to simultaneously solve a set of equations: when determining whether a set was linearly independent, when determining if a vector was in the span of a set, and when determining the formula for the cross product. For each of these cases, these examples could be solved, with some effort, using high school methods of substitution and elimination. However, in the real world, we don't often get a set of 3 linear equations in 3 unknowns. Problems can have hundreds of thousands of equations and hundreds of thousands of unknowns. Hence, we want to develop some theory for solving equations that will allow us to solve such problems.

It is important to remember while studying this chapter that in typical applications the computations are performed by computer. Thus, although we expect that you can solve small systems by hand, it is more important that you understand how to set up an appropriate system of equations and to interpret the results. In particular, the theory presented in this chapter will be extremely important throughout Math 136 and Math 235.

2.1 Systems of Linear Equations

DEFINITION
Linear Equation

An equation of the form

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

where $a_1, \ldots, a_n, b$ are constants is called a **linear equation**. The constants a_i are called the **coefficients** of the equation and b is called the **right-hand side**.

REMARKS

1. In this course the coefficients and right-hand side will typically be real numbers; however, we will occasionally present linear equations where the coefficients and right-hand side are complex.

2. From our work in the last chapter, we know that a linear equation where $a_1, \ldots, a_n, b \in \mathbb{R}$ in n variables geometrically represents a hyperplane in $\mathbb{R}^n$.

EXAMPLE 1

The equation $x_1 + 2x_3 = 4$ is linear. Note that the coefficient of x_2 in this equation is 0.

The equation $x_2 = -4x_1$ is also linear as it can be written as $4x_1 + x_2 = 0$ (called **standard form**).

The equations $x_1^2 + x_1x_2 = 5$ and $x \sin x = 0$ are not linear equations.

DEFINITION

System of Linear Equations

A set of m linear equations in the same variables $x_1, \ldots, x_n$ is called a **system of linear equations**.

A general system of m linear equations in n variables has the form

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

Observe that the coefficient a_{ij} represents the coefficient of x_j in the i-th equation.

DEFINITION

Solution of a System

A vector $\vec{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} \in \mathbb{R}^n$ is called a **solution** of a system of m linear equations in n unknowns (variables) if all m equations are satisfied when we set $x_i = s_i$ for $1 \leq i \leq n$.

DEFINITION

Consistent

Inconsistent

If a system of linear equations has at least one solution, then it is said to be **consistent**. Otherwise, it is said to be **inconsistent**.

EXAMPLE 2

A solution of the system of 3 linear equations in 2 unknowns

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 - 3x_2 &= 12 \\ -2x_1 - 3x_2 &= 0 \end{aligned}$$

is $\vec{s} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$, since if we take $x_1 = 3$ and $x_2 = -2$, then we get

$$\begin{aligned} 3 + (-2) &= 1 \\ 2(3) - 3(-2) &= 12 \\ -2(3) - 3(-2) &= 0 \end{aligned}$$

Therefore, this system is consistent.

EXAMPLE 3 Two of the solutions of the system of 2 equations in 3 unknows

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

are $\vec{s}_1 = \begin{bmatrix} 7 \\ 2 \\ 0 \end{bmatrix}$ and $\vec{s}_2 = \begin{bmatrix} 5 \\ 1 \\ 1 \end{bmatrix}$. Therefore, this system is consistent.

EXAMPLE 4 The system of 3 equations in 3 unknowns

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

does not have any solutions. Hence, the system is inconsistent.

REMARK

Observe that geometrically a system of m linear equations in n variables represents m hyperplanes in $\mathbb{R}^n$ and so a solution of the system is a vector in $\mathbb{R}^n$ which lies on all m hyperplanes. The system is inconsistent if all m hyperplanes do not share a point of intersection.

It is easy to show geometrically that a system of linear equations in 2 variables is either inconsistent, consistent with a unique solution, or consistent with infinitely many solutions (you should sketch all 3 possibilities). Of course, it is clear that we can make a system of linear equations that is inconsistent, or that is consistent with a unique solution. We now prove that if a system of equations is consistent with more than one solution, then it must in fact have infinitely many solutions.

THEOREM 1

Assume the system of linear equations with $a_1, \ldots, a_n, b \in \mathbb{R}$

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

has two distinct solutions $\vec{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$ and $\vec{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$. Then, $\vec{x} = \vec{s} + c(\vec{s} - \vec{t})$ is a distinct solution for each $c \in \mathbb{R}$.

Proof: For each i, $1 \leq i \leq m$ we have

$$\begin{aligned} a_{i1}\big(s_1 + c(s_1 - t_1)\big) + \cdots + a_{in}\big(s_n + c(s_n - t_n)\big) \\ &= a_{i1}s_1 + ca_{i1}s_1 - ca_{i1}t_1 + \cdots + a_{in}s_n + ca_{in}s_n - ca_{in}t_n \\ &= a_{i1}s_1 + \cdots + a_{in}s_n + c\big(a_{i1}s_1 + \cdots + a_{in}s_n\big) + c\big(a_{i1}t_1 + \cdots + a_{in}t_n\big) \\ &= b_i + cb_i - cb_i \\ &= b_i \end{aligned}$$

Thus, $\vec{s} + c(\vec{s} - \vec{t})$ is a solution for each $c \in \mathbb{R}$.

We now show that for any $c_1 \neq c_2$, the solutions $\vec{s} + c_1(\vec{s} - \vec{t})$ and $\vec{s} + c_2(\vec{s} - \vec{t})$ are distinct. Assume that they are not distinct. Then

$$\begin{aligned} \vec{s} + c_1(\vec{s} - \vec{t}) &= \vec{s} + c_2(\vec{s} - \vec{t}) \\ c_1(\vec{s} - \vec{t}) &= c_2(\vec{s} - \vec{t}) \\ (c_1 - c_2)(\vec{s} - \vec{t}) &= \vec{0} \end{aligned}$$

Since $c_1 \neq c_2$, this implies that $\vec{s} - \vec{t} = \vec{0}$ and hence $\vec{s} = \vec{t}$. But this contradicts our assumption that $\vec{s} \neq \vec{t}$. $\square$

DEFINITION

Solution Set

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

In the next section, we will derive a couple of methods for finding the solution set of a system of linear equations.

2.2 Solving Systems of Linear Equation

Given a system of 2 linear equations in 2 unknowns, we can use high school methods of substitution and elimination to solve it.

EXAMPLE 1 Solve the system

$$\begin{aligned} 2x_1 + 3x_2 &= 11 \\ 3x_1 + 6x_2 &= 7 \end{aligned}$$

Solution: We first want to solve for one variable, say x_2. To do this, we need to eliminate x_1 from one of the equations. If we multiply the first equation by (-3) the system becomes

$$\begin{aligned} -6x_1 - 9x_2 &= -33 \\ 3x_1 + 6x_2 &= 7 \end{aligned}$$

Now, if we multiply the second equation by 2 we get

$$\begin{aligned} -6x_1 - 9x_2 &= -33 \\ 6x_1 + 12x_2 &= 14 \end{aligned}$$

Adding the first equation to the second we get

$$\begin{aligned} -6x_1 - 9x_2 &= -33 \\ 0x_1 + 3x_2 &= -19 \end{aligned}$$

Multiplying the second equation by $\frac{1}{3}$ gives

$$\begin{aligned} -6x_1 - 9x_2 &= -33 \\ 0x_1 + x_2 &= -19/3 \end{aligned}$$

We can now add 9 times the second equation to the first equation to get

$$\begin{aligned} -6x_1 + 0x_2 &= -90 \\ 0x_1 + x_2 &= -19/3 \end{aligned}$$

Finally, multiplying the top equation by $-\frac{1}{6}$ we get

$$\begin{aligned} x_1 + 0x_2 &= 15 \\ 0x_1 + x_2 &= -19/3 \end{aligned}$$

Hence, the system is consistent with unique solution $\begin{bmatrix} 15 \\ -19/3 \end{bmatrix}$.

Observe that after each step we actually have a new system of linear equations to solve. Of course, the idea is that each new system has the same solution set as the original and is easier to solve.

DEFINITION

Equivalent

Two systems of equations are said to be **equivalent** if they have the same solution set.

Of course, this method can be used to solve larger systems as well. To invent a method for solving very large systems, it is important that you really understand how the method works for smaller systems.

EXERCISE 1

Solve the system by using substitution and elimination. Clearly show/explain all of your steps used in solving the system and describe your general procedure.

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

Now imagine that you need to solve a system of a hundred thousand equations with a hundred thousand variables. Certainly you are not going to want to do this by hand (it would be environmentally unfriendly to use that much paper). So, being clever, you decide that you want to program a computer to solve the system for you. To do this, you need to ask yourself a few questions: How is the computer going to store the system? What operations is the computer allowed to perform on the equations? How will the computer know which operations it should use to solve the system?

To answer the first question, we observe that when using substitution and elimination, we do not really need to write down the variables x_i each time as we are only modifying the coefficients and right-hand side. Thus, it makes sense to store just the coefficients and right-hand side of each equation in the system in a rectangular array.

DEFINITION

Augmented Matrix
Coefficient Matrix

The **augmented matrix** for the system of linear equations

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

is the rectangular array

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

The **coefficient matrix** of the system is

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

EXAMPLE 2

The system of linear equations

$$\begin{aligned} 3x_1 + x_2 + x_3 &= -1 \\ 6x_1 \qquad + x_3 &= -2 \\ -3x_1 + 2x_2 &= 3 \end{aligned}$$

has coefficient matrix

$$\begin{bmatrix} 3 & 1 & 1 \\ 6 & 0 & 1 \\ -3 & 2 & 0 \end{bmatrix}$$

and augmented matrix

$$\left[\begin{array}{ccc|c} 3 & 1 & 1 & -1 \\ 6 & 0 & 1 & -2 \\ -3 & 2 & 0 & 3 \end{array}\right]$$

REMARKS

1. It is very important to observe that we can look at the coefficient matrix in two ways. First, the i-th row of the coefficient matrix represents the coefficients of the i-th equation in the system. Second, the j-th column of the coefficient matrix represents the coefficients of x_j in all equations.
2. We will denote a system of linear equations with coefficient matrix A and right-hand side $\vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ by $\left[A \mid \vec{b} \right]$. In the next chapter, we will introduce another way of denoting a system of equations.

To answer the question about what operations is the computer allowed to perform, it is helpful to write each step of our example in matrix form.

EXAMPLE 3

Write each step of Example 1 in matrix form and describe how the rows of the matrix were modified in each step.

Solution: The initial system was

$$\left[\begin{array}{cc|c} 2 & 3 & 11 \\ 3 & 6 & 7 \end{array}\right]$$

We then multiplied the first row of the augmented matrix by (-3) to get the augmented matrix

$$\left[\begin{array}{cc|c} -6 & -9 & -33 \\ 3 & 6 & 7 \end{array}\right]$$

We then multiplied the second row by 2 to get

$$\left[\begin{array}{cc|c} -6 & -9 & -33 \\ 6 & 12 & 14 \end{array}\right]$$

Adding the first row to the second row gave

$$\left[\begin{array}{cc|c} -6 & -9 & -33 \\ 0 & 3 & -19 \end{array}\right]$$

Next we multiplied the second row by $\frac{1}{3}$ to get

$$\left[\begin{array}{cc|c} -6 & -9 & -33 \\ 0 & 1 & -19/3 \end{array}\right]$$

Then we added 9 times the second row to the first.

$$\left[\begin{array}{cc|c} -6 & 0 & -90 \\ 0 & 1 & -19/3 \end{array}\right]$$

The last step was to multiply the first row by $-\frac{1}{6}$.

$$\left[\begin{array}{cc|c} 1 & 0 & 15 \\ 0 & 1 & -19/3 \end{array}\right]$$

This is the augmented matrix for the system

$$\begin{aligned} x_1 &= 15 \\ x_2 &= -19/3 \end{aligned}$$

which gives us the solution.

EXERCISE 2

Write each step of Exercise 1 in matrix form and describe how the rows of the matrix were modified in each step.

Our method of solving has involved two types of operations. One was to multiply a row by a non-zero number and the second was to add a multiple of one row to another. However, to create a nice method for a computer to solve a system, we will add one more type of operation: swapping rows.

DEFINITION

Elementary Row Operations

The three **elementary row operations** (EROs) for a solving a system of linear equations are:

1. multiplying a row by a non-zero scalar,
2. adding a multiple of one row to another,
3. swap two rows.

When applying EROs to a matrix, called **row reducing the matrix**, we often use short hand notation to indicate which operations we are using:

1. cR_i indicates multiplying the i-th row by $c \neq 0$.
2. $R_i + cR_j$ indicates adding c times the j-th row to the i-th row.
3. $R_i \leftrightarrow R_j$ indicates swapping the i-th row and the j-th row.

For now you are required to always indicate which EROs you use in each step. Not only will this help you in row reducing a matrix, and finding/fixing computational errors, but later in the course, we will find a case where this is necessary.

Of course, our goal is to row reduce the augmented matrix of a system into the augmented matrix of an equivalent system that is easier to solve.

DEFINITION

Row Equivalent

Two matrices A and B are said to be **row equivalent** if there exists a sequence of elementary row operations that transform A into B.

REMARK

Observe that EROs are reversible (we can always perform an ERO that undoes what an ERO did), hence if there exists a sequence of elementary row operations that transform A into B, then there also exists a sequence of elementary row operations that transforms B into A.

THEOREM 1

If the augmented matrices $\left[A_1 \mid \vec{b}_1\right]$ and $\left[A \mid \vec{b}\right]$ are row equivalent, then the systems of linear equations associated with each system are equivalent.

Proof: The proof is left as an exercise.

Finally, to answer the third question about which operations a computer should use to solve the system, we need to figure out an algorithm for the computer to follow. The algorithm should apply elementary row operations to the augmented matrix of a system until we have an equivalent system which is easy to solve.

DEFINITION

Reduced Row Echelon Form

A matrix R is said to be in **reduced row echelon form** (RREF) if:

1. All rows containing a non-zero entry are above rows which only contains zeros.
2. The first non-zero entry in each non-zero row is 1, called a **leading one**.
3. The leading one in each non-zero row is to the right of the leading one in any row above it.
4. A leading one is the only non-zero entry in its column.

If R is row equivalent to a matrix A, then we say that R is the reduced row echelon form of A.

REMARK

The reduced row echelon form is sometimes called the row canonical form.

EXAMPLE 4

Determine which of the following matrices are in RREF. a) $\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ b) $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ c) $\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ d) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

Solution: The matrices in a) and b) are in RREF. The matrix in c) is not in RREF as the column containing the leading one in the third row has another non-zero entry. The matrix d) is not in RREF since the leading one in the second row is to the left of the leading one in the row above it.

THEOREM 2

The RREF of a matrix is unique.

Proof: The proof requires some additional tools which will be presented in Chapter 4.

EXAMPLE 5

Solve the following system of equations by row reducing the augmented matrix of the system to RREF.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & & & = & -7 \\ 2x_1 & + & 4x_2 & + & x_3 & = & -16 \\ x_1 & + & 2x_2 & + & x_3 & = & 9 \end{array}$$

Solution: We get

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & -7 \\ 2 & 4 & 1 & -16 \\ 1 & 2 & 1 & 9 \end{array}\right] \begin{array}{l} \\ R_2 - 2R_1 \\ R_3 - R_1 \end{array} \sim \left[\begin{array}{ccc|c} 1 & 1 & 0 & -7 \\ 0 & 2 & 1 & -2 \\ 0 & 1 & 1 & 16 \end{array}\right] \begin{array}{l} \\ \\ R_2 \leftrightarrow R_3 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & -7 \\ 0 & 1 & 1 & 16 \\ 0 & 2 & 1 & -2 \end{array}\right] \begin{array}{l} R_1 - R_2 \\ \\ R_3 - 2R_2 \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & -23 \\ 0 & 1 & 1 & 16 \\ 0 & 0 & -1 & -34 \end{array}\right] \begin{array}{l} \\ \\ (-1)R_3 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & -23 \\ 0 & 1 & 1 & 16 \\ 0 & 0 & 1 & 34 \end{array}\right] \begin{array}{l} R_1 + R_2 \\ R_2 - R_3 \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 11 \\ 0 & 1 & 0 & -18 \\ 0 & 0 & 1 & 34 \end{array}\right]$$

Hence the solution is $\vec{x} = \begin{bmatrix} 11 \\ -18 \\ 34 \end{bmatrix}$.

EXAMPLE 6 Find the solution set of the following system of equations.

$$\begin{array}{ccccccccc} x_1 & + & x_2 & + & 2x_3 & + & x_4 & = & 3 \\ x_1 & + & 2x_2 & + & 4x_3 & + & x_4 & = & 7 \\ x_1 & & & & & + & x_4 & = & -21 \end{array}$$

Solution: We write the coefficient matrix and row reduce the coefficient matrix to RREF.

$$\left[\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 3 \\ 1 & 2 & 4 & 1 & 7 \\ 1 & 0 & 0 & 1 & -21 \end{array}\right] \begin{array}{l} \\ R_2 - R_1 \\ R_3 - R_1 \end{array} \sim \left[\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 3 \\ 0 & 1 & 2 & 0 & 4 \\ 0 & -1 & -2 & 0 & -24 \end{array}\right] \begin{array}{l} R_1 - R_2 \\ \\ R_3 + R_2 \end{array} \sim$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 & -20 \end{array}\right]$$

Observe that the last row represents the equation $0x_1 + 0x_2 + 0x_3 + 0x_4 = -20$. Clearly this is impossible, so the system is inconsistent.

It is clear that if we ever get a row of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & b \end{array}\right]$ with $b \neq 0$, then the system of equations inconsistent as this corresponds to the equation $0 = b$.

Infinitely Many Solutions

As we discussed above, a consistent system of linear equations either has a unique solution or infinitely many. We now look at how to determine all solutions of a system with infinitely many solutions. We start by considering an example.

EXAMPLE 7

Show the following system of equations has infinitely many solutions and find the solution set.

$$\begin{aligned} x_1 + x_2 + x_3 &= 4 \\ x_2 + x_3 &= 3 \end{aligned}$$

Solution: We row reduce the augmented matrix to RREF.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & 1 & 3 \end{array}\right] \begin{array}{c} R_1 - R_2 \\ \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 3 \end{array}\right]$$

Observe that the RREF of the augmented matrix corresponds to the system of equations

$$\begin{aligned} x_1 &= 1 \\ x_2 + x_3 &= 3 \end{aligned}$$

There are infinitely many choices for x_2 and x_3 that will satisfy $x_2 + x_3 = 3$. Hence, the system has infinitely many solutions.

Observe that by choosing any $x_3 = t \in \mathbb{R}$, we get a unique value for x_2; in particular, $x_2 = 3 - t$. Thus, the solution set of the system is

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3-t \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

for any $t \in \mathbb{R}$.

REMARK

The geometric interpretation of the example above is that the planes $x_1 + x_2 + x_3 = 4$ and $x_2 + x_3 = 3$ in $\mathbb{R}^3$ intersect in the line $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, $t \in \mathbb{R}$.

DEFINITION
Free Variable

Let R be the RREF of a coefficient matrix of a system of linear equations. If the j-th column of R does not contain a leading one, then we call x_j a **free variable**.

Using the examples above, we can give an algorithm for solving a system of linear equations.

ALGORITHM

To solve a system of linear equations:

1. Write the augmented matrix for the system.
2. Use elementary row operations to row reduce the augmented matrix into RREF.
3. Write the system of linear equations corresponding to the RREF.
4. If the system contains an equation of the form $0 = b$ where $b \neq 0$, then stop as the system is inconsistent.
5. Otherwise, move each free variable (if any) to the right hand side of each equation and assign each free variable a parameter.
6. Determine the solution set by using vector operations to write the system as a linear combination of vectors.

REMARK

The typical method for row reducing an augmented matrix to its RREF is called Gauss-Jordan Elimination. However, although this algorithm for row reducing a matrix always works, it is not always the most efficient way to reduce a matrix by hand (or by computer). In individual cases, you may want to perform various "tricks" to make the row reductions/computations easier. For example, you probably want to avoid fractions as long as possible. These "tricks" are best learned from experience. That is, you should row reduce lots of matrices.

EXAMPLE 8 Solve the following system of linear equations.

$$\begin{aligned} 2x_1 + x_2 + x_3 + x_4 &= -2 \\ x_1 + x_3 + x_4 &= 1 \\ -2x_2 + 2x_3 + 2x_4 &= 8 \end{aligned}$$

Solution: We have

$$\left[\begin{array}{cccc|c} 2 & 1 & 1 & 1 & -2 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & -2 & 2 & 2 & 8 \end{array}\right] \begin{array}{c} R_1 \leftrightarrow R_2 \\ \\ \frac{1}{2}R_3 \end{array} \sim \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & -2 \\ 0 & -1 & 1 & 1 & 4 \end{array}\right] \; R_2 - 2R_1 \; \sim$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & -1 & -4 \\ 0 & -1 & 1 & 1 & 4 \end{array}\right] \begin{array}{c} \\ \\ R_3 + R_2 \end{array} \sim \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & -1 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The corresponding system of equations is

$$\begin{aligned} x_1 + x_3 + x_4 &= 1 \\ x_2 - x_3 - x_4 &= -4 \\ 0 &= 0 \end{aligned}$$

Since x_3 and x_4 are free variables we let $x_3 = s \in \mathbb{R}$ and $x_4 = t \in \mathbb{R}$. Thus, we have $x_1 = 1 - x_3 - x_4 = 1 - s - t$ and $x_2 = -4 + x_3 + x_4 = -4 + s + t$. Therefore all solutions have the form

$$\begin{aligned} \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} &= \begin{bmatrix} 1 - s - t \\ -4 + s + t \\ s \\ t \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -4 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

We can solve systems of linear equations with complex coefficients over the complex numbers $\mathbb{C}$ using exactly the same method.

EXAMPLE 9 Solve the following system over $\mathbb{C}$.

$$\begin{array}{rcrcrcr} & & z_2 & - & iz_3 & = & 1 + 3i \\ iz_1 & - & z_2 & + & (-1+i)z_3 & = & 1 + 2i \\ 2z_1 & + & 2iz_2 + & & (3+2i)z_3 & = & 4 \end{array}$$

Solution: We have

$$\left[\begin{array}{ccc|c} 0 & 1 & -i & 1+3i \\ i & -1 & -1+i & 1+2i \\ 2 & 2i & 3+2i & 4 \end{array}\right] \begin{array}{c} R_1 \leftrightarrow R_2 \end{array} \sim \left[\begin{array}{ccc|c} i & -1 & -1+i & 1+2i \\ 0 & 1 & -i & 1+3i \\ 2 & 2i & 3+2i & 4 \end{array}\right] \begin{array}{c} (-i)R_1 \\ \\ \\ \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & i & 1+i & 2-i \\ 0 & 1 & -i & 1+3i \\ 2 & 2i & 3+2i & 4 \end{array}\right] \begin{array}{c} \\ \\ R_3 - 2R_1 \end{array} \sim \left[\begin{array}{ccc|c} 1 & i & 1+i & 2-i \\ 0 & 1 & -i & 1+3i \\ 0 & 0 & 1 & 2i \end{array}\right] \begin{array}{c} R_1 - iR_2 \\ \\ \\ \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 0 & i & 5-2i \\ 0 & 1 & -i & 1+3i \\ 0 & 0 & 1 & 2i \end{array}\right] \begin{array}{c} R_1 - iR_3 \\ R_2 + iR_3 \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 7-2i \\ 0 & 1 & 0 & -1+3i \\ 0 & 0 & 1 & 2i \end{array}\right]$$

EXAMPLE 10

Solve the following system over $\mathbb{C}$.

$$\begin{aligned}(1+i)z_1 + 2iz_2 - iz_3 &= i\\ 2iz_2 + z_3 &= i\\ z_1 - z_3 &= 0\end{aligned}$$

Solution:

$$\left[\begin{array}{ccc|c} 1+i & 2i & -i & i\\ 0 & 2i & 1 & i\\ 1 & 0 & -1 & 0\end{array}\right] \begin{array}{c} R_1 \leftrightarrow R_3 \\ \\ \\ \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0\\ 0 & 2i & 1 & i\\ 1+i & 2i & -i & i\end{array}\right] \begin{array}{c} \\ \frac{-i}{2}R_2 \\ R_3-(1+i)R_1 \end{array} \sim$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0\\ 0 & 1 & -i/2 & 1/2\\ 0 & 2i & 1 & i\end{array}\right] \begin{array}{c} \\ \\ R_3-2iR_2 \end{array} \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0\\ 0 & 1 & -i/2 & 1/2\\ 0 & 0 & 0 & 0\end{array}\right]$$

The corresponding system of equations is

$$\begin{aligned}z_1 - z_3 &= 0\\ z_2 - \frac{i}{2}z_3 &= \frac{1}{2}\end{aligned}$$

Since z_3 is a complex free variable we let $z_3 = t \in \mathbb{C}$. Thus, we get $z_1 = t$ and $z_2 = \frac{1}{2} + \frac{i}{2}t$, hence the solution is

$$\vec{z} = \begin{bmatrix} z_1\\ z_2\\ z_3\end{bmatrix} = \begin{bmatrix} t\\ \frac{1}{2}+\frac{i}{2}t\\ t\end{bmatrix} = \begin{bmatrix} 0\\ 1/2\\ 0\end{bmatrix} + t\begin{bmatrix} 1\\ i/2\\ 1\end{bmatrix}$$

Rank

It is clear that the number of free variables in a system of linear equations is the number of columns in the RREF of the coefficient matrix which do not have leading ones. It is therefore not surprising that counting the number of leading ones can be very useful.

DEFINITION
Rank

The **rank** of a matrix is the number of leading ones in the RREF of the matrix.

EXAMPLE 11

Determine the rank of the following matrices.

a) $A = \begin{bmatrix} 1 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & -2\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$.

Solution: Since A is in RREF, we see that A has two leading ones, so $\operatorname{rank} A = 2$.

b) $B = \begin{bmatrix} 2 & 1 & 1 & 1 & -2 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 & 8 \end{bmatrix}$.

Solution: Observe that the matrix B is just the augmented matrix in Example 8. Hence, performing the same row operations we find that the RREF of B is

$$R = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & -1 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore, $\operatorname{rank} B = 2$.

We now get the following extremely important theorem which summarizes all of our results on system of linear equations in terms of rank. This theorem will be referenced many times throughout the rest of these notes.

THEOREM 3 Let A be the $m \times n$ coefficient matrix of a system of linear equations.

(1) If the rank of A is less than the rank of the augmented matrix $\left[A \mid \vec{b}\right]$, then the system is inconsistent.

(2) If the system $\left[A \mid \vec{b}\right]$ is consistent, then the system contains $n - \operatorname{rank} A$ free variables (parameters). In particular, a consistent system has a unique solution if and only if $\operatorname{rank} A = n$.

(3) $\operatorname{rank} A = m$ if and only if the system $\left[A \mid \vec{b}\right]$ is consistent for every $\vec{b} \in \mathbb{R}^m$.

Proof: (1) If the rank of A is less than the rank of the augmented matrix, then there is a row in the RREF of the augmented matrix of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & 1 \end{array}\right]$. This corresponds to the equation $0 = 1$, which implies the system is inconsistent.

(2) This follows immediately from the definition of rank and free variables.

(3) Assume that the RREF of $\left[A \mid \vec{b}\right]$ is $[R \mid \vec{c}]$. If $\left[A \mid \vec{b}\right]$ is inconsistent for some $\vec{b} \in \mathbb{R}^m$, then the RREF of the augmented matrix must contain a row of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & 1 \end{array}\right]$, as otherwise, we could find a solution to $\left[A \mid \vec{b}\right]$ by picking all free-variables to be 0 and setting c_i equal to the variable contains the leading one in the i-th column. Therefore, $\operatorname{rank} A < m$.

On the other hand, if $\operatorname{rank} A < m$, then the RREF R of A has a row of all zeros. Thus, $[R \mid \vec{e}_n]$ is inconsistent as it has a row of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & 1 \end{array}\right]$. Since elementary row operations are reversible, we can apply the reverse operations needed to row reduce A to R on $[R \mid \vec{e}_n]$ to get $\left[A \mid \vec{b}\right]$ for some $\vec{b} \in \mathbb{R}^n$. Then, this system is inconsistent since elementary row operations do not change the solution set. Thus, there exists some $\vec{b} \in \mathbb{R}^m$ such that $\left[A \mid \vec{b}\right]$ is inconsistent. $\square$

Homogeneous Systems

DEFINITION
Homogeneous System

A system of linear equations is said to be a **homogeneous system** if the right-hand side only contains zeros. That is, it has the form $\left[A \mid \vec{0}\right]$.

We will soon see that homogeneous systems arise frequently in linear algebra. Of course, all of our theory above regarding systems of linear equations applies to homogeneous systems. But, the fact that the right-hand side only contains 0s simplifies a few things.

Observe that the last column of any matrix that is row equivalent to the augmented matrix of a homogeneous system will only contain zeros. Thus, it is standard practice to omit this row when solving a homogeneous system; we just row reduce the coefficient matrix. From this, we immediately observe that it is impossible to get a row of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & 1 \end{array}\right]$, hence a homogeneous system is always consistent. In particular, it is easy to see that the zero vector $\vec{0}$ is always a solution of a homogeneous system. This is called the **trivial solution**.

EXAMPLE 12

Find the solution set of the homogeneous system

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 0 \\ -2x_1 - 3x_2 - 6x_3 &= 0 \\ x_1 + 3x_2 + 4x_3 &= 0 \\ x_1 + 4x_2 + 5x_3 &= 0 \end{aligned}$$

Solution: We row reduce the coefficient matrix to get

$$\begin{bmatrix} 1 & 2 & 3 \\ -2 & -3 & -6 \\ 1 & 3 & 4 \\ 1 & 4 & 5 \end{bmatrix} \begin{matrix} \\ R_2 + 2R_1 \\ R_3 - R_1 \\ R_4 - R_1 \end{matrix} \sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix} \begin{matrix} R_1 - 2R_2 \\ \\ R_3 - R_2 \\ R_4 - 2R_2 \end{matrix} \sim$$

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix} \begin{matrix} R_1 - 3R_3 \\ \\ \\ R_4 - 2R_3 \end{matrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

This gives $x_1 = 0$, $x_2 = 0$, and $x_3 = 0$. Therefore, the only solution is the trivial solution.

EXAMPLE 13

Find the solution set of the homogeneous system

$$\begin{aligned} x_1 + 2x_2 &= 0 \\ x_1 + 3x_2 + x_3 &= 0 \\ x_2 + x_3 &= 0 \end{aligned}$$

Solution: We row reduce the coefficient matrix to get

$$\begin{bmatrix} 1 & 2 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{matrix} \\ R_2 - R_1 \end{matrix} \sim \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{matrix} R_1 - 2R_2 \\ R_3 - R_2 \end{matrix} \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

The corresponding system is

$$\begin{aligned} x_1 - 2x_3 &= 0 \\ x_2 + x_3 &= 0 \end{aligned}$$

Since x_3 is a free variable, we let $x_3 = t \in \mathbb{R}$. Thus, the solution set has vector equation

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

We see in both of the examples above that we could have written the solution set as a spanning set. Which, by Theorem 2, indicates that both solution sets are subspaces of $\mathbb{R}^3$.

THEOREM 4

The solution set of a homogeneous system of m linear equations in n variables is a subspace of $\mathbb{R}^n$.

Proof: The proof is left as an exercise.

REMARK

Because the solution set of a homogeneous system is a subspace, we often call it the **solution space** of the system.

Chapter 3

Matrices and Linear Mappings

3.1 Operations on Matrices

Addition and Scalar Multiplication of Matrices

In the last chapter we used matrix notation to help us solve systems of linear equations. However, matrices can also be used in other settings as well. In particular, we can treat matrices like vectors.

DEFINITION
Matrix

An $m \times n$ **matrix** A is a rectangular array with m rows and n columns. We denote the entry in the i-th row and j-th column by a_{ij}. That is,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}$$

REMARKS

1. Two matrices A and B are **equal** if and only if they have the same size and have corresponding entries equal. That is, if $a_{ij} = b_{ij}$ for $1 \le i \le m$, $1 \le j \le n$.
2. When working with multiple matrices we sometimes denote the ij-th entry of a matrix A by $(A)_{ij} = a_{ij}$.
3. The set of all $m \times n$ matrices with real entries is denoted $M_{m\times n}(\mathbb{R})$.

We define addition and scalar multiplication of matrices to match what we did with vectors in $\mathbb{R}^n$.

DEFINITION

Addition and Scalar Multiplication

Let $A, B \in M_{m\times n}(\mathbb{R})$ and $c \in \mathbb{R}$. Then we define $A+B$ and cA by

$$(A+B)_{ij} = (A)_{ij} + (B)_{ij}$$
$$(cA)_{ij} = c(A)_{ij}$$

EXAMPLE 1

a) $\begin{bmatrix} 1 & 2 \\ 3 & -3 \end{bmatrix} + \begin{bmatrix} 2 & 5 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 7 \\ 2 & 0 \end{bmatrix}$.

b) $\sqrt{2}\begin{bmatrix} 1 & -2 \\ -\sqrt{2} & \sqrt{3} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & -2\sqrt{2} \\ -2 & \sqrt{6} \end{bmatrix}$.

c) $3\begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} - 2\begin{bmatrix} 2 & 2 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 3-4 & -3-4 \\ 6+4 & 12-6 \end{bmatrix} = \begin{bmatrix} -1 & -7 \\ 10 & 6 \end{bmatrix}$.

As we did with vectors in $\mathbb{R}^n$, we will call a sum of scalar multiples of matrices a linear combination. Note that this only makes sense if all of the matrices are of the same size.

Properties of Matrix Addition and Multiplication by Scalars

THEOREM 1

Let A, B, and C be $m \times n$ matrices and let c_1, c_2 be real scalars. Then

V1 $A+B$ is an $m \times n$ matrix;
V2 $(A+B)+C = A+(B+C)$;
V3 $A+B = B+A$;
V4 There exists a matrix, denoted by $O_{m,n}$, such that $A+O_{m,n} = A$. In particular, $O_{m,n}$ is the $m \times n$ matrix with all entries zero and is called the **zero matrix**.
V5 For each matrix A, there exists an $m \times n$ matrix $(-A)$, with the property that $A+(-A) = O_{m,n}$. In particular, $(-A)$ is defined by $(-A)_{ij} = -(A)_{ij}$;
V6 c_1A is an $m \times n$ matrix;
V7 $c_1(c_2A) = (c_1c_2)A$;
V8 $(c_1+c_2)A = c_1A + c_2A$;
V9 $c_1(A+B) = c_1A + c_1B$;
V10 $1A = A$.

Proof: These properties follow easily from the definitions of addition and multiplication by scalars. The proof is left to the reader.

REMARK

Observe that these are exactly the same ten properties we had in Section 1.1 for addition and multiplication by scalars of vectors in $\mathbb{R}^n$.

The Transpose of a Matrix

As we will soon see, we will sometimes wish to work with the rows of an $m \times n$ matrix as vectors in $\mathbb{R}^n$. To preserve our convention of writing vectors in $\mathbb{R}^n$ as column vectors, we invent some notation for turning rows into columns and vice versa.

DEFINITION

Transpose

The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix whose ij-th entry is the ji-th entry of A. We denote the transpose of A by A^T. Hence, we have

$$(A^T)_{ij} = (A)_{ji}$$

EXAMPLE 2

Determine the transpose of $A = \begin{bmatrix} 1 & 2 & 5 \end{bmatrix}$, $B = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$, and $C = \begin{bmatrix} 3 & -2 \\ 4 & 1 \\ 7 & -6 \end{bmatrix}$.

Solution: We have

$$A^T = \begin{bmatrix} 1 & 2 & 5 \end{bmatrix}^T = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

$$B^T = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}^T = \begin{bmatrix} -3 & 0 & 1 \end{bmatrix}$$

$$C^T = \begin{bmatrix} 3 & -2 \\ 4 & 1 \\ 7 & -6 \end{bmatrix}^T = \begin{bmatrix} 3 & 4 & 7 \\ -2 & 1 & -6 \end{bmatrix}$$

THEOREM 2

For any $m \times n$ matrices A and B and scalar $c \in \mathbb{R}$ we have

(1) $(A^T)^T = A$.

(2) $(A + B)^T = A^T + B^T$.

(3) $(cA)^T = cA^T$.

Proof: (1) By definition of the transpose we have $\left((A^T)^T\right)_{ij} = (A^T)_{ji} = A_{ij}$.

(2) $\left((A+B)^T\right)_{ij} = (A+B)_{ji} = (A)_{ji} + (B)_{ji} = (A^T)_{ij} + (B^T)_{ij} = (A^T + B^T)_{ij}$.

(3) $\left((cA)^T\right)_{ij} = (cA)_{ji} = c(A)_{ji} = c(A^T)_{ij} = (cA^T)_{ij}$. □

As mentioned above, one of the main use of the transpose is so that we have notation for the rows of a matrix. We demonstrate this in the next example.

EXAMPLE 3

Let $\vec{a}_1 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ and $\vec{a}_2 = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$. Then the matrix $A = \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \end{bmatrix}$ is $\begin{bmatrix} 1 & 3 & 1 \\ 2 & -1 & 4 \end{bmatrix}$.

EXERCISE 1

If $\begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \vec{a}_3^T \end{bmatrix} = \begin{bmatrix} 1 & 3 & \\ 1 & & \\ -4 & 0 & 2 \\ 5 & 9 & -3 \end{bmatrix}$, then what is $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$?

Matrix-Vector Multiplication

We saw in the last chapter that we had two ways of viewing the coefficient matrix of A of a system of linear equations $\left[A \mid \vec{b}\right]$: the coefficients of the i-th equation are the entries in the i-th row of A, or the coefficients of x_i are the entries in the i-th column of A. We will now use both of these interpretations to derive equivalent definitions of matrix-vector multiplication so that we can represent the system $\left[A \mid \vec{b}\right]$ by the equation $A\vec{x} = \vec{b}$ where $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Note that both interpretations are going to be very useful throughout this course and in Linear Algebra 2.

Coefficients are the Rows of A

Let A be an $m \times n$ matrix. Using the transpose, we can write $A = \begin{bmatrix} \vec{a}_1^T \\ \vdots \\ \vec{a}_m^T \end{bmatrix}$ where $\vec{a}_i \in \mathbb{R}^n$. Observe that the i-th equation of the system $\left[A \mid \vec{b}\right]$ can be written $\vec{a}_i \cdot \vec{x} = b_i$ since

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + \cdots + a_nx_n$$

Therefore, the system $\left[A \mid \vec{b}\right]$ can be written as

$$\begin{bmatrix} \vec{a}_1 \cdot \vec{x} \\ \vdots \\ \vec{a}_m \cdot \vec{x} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

To write the left-hand side in the form $A\vec{x}$, we make the following definition.

DEFINITION

Matrix-Vector Multiplication

Let A be an $m \times n$ matrix whose rows are denoted $\vec{a}_i^T$ for $1 \leq i \leq m$. Then, for any $\vec{x} \in \mathbb{R}^n$, we define

$$A\vec{x} = \begin{bmatrix} \vec{a}_1 \cdot \vec{x} \\ \vdots \\ \vec{a}_m \cdot \vec{x} \end{bmatrix}$$

REMARK

If A is an $m \times n$ matrix, then $A\vec{x}$ is only defined if $\vec{x} \in \mathbb{R}^n$. Moreover, if $\vec{x} \in \mathbb{R}^n$, then $A\vec{x} \in \mathbb{R}^m$.

EXAMPLE 4

Calculate $\begin{bmatrix} 1 & 3 \\ 2 & -4 \\ 9 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Solution: Let $\begin{bmatrix} 1 & 3 \\ 2 & -4 \\ 9 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Then, we have

$$\begin{aligned} b_1 &= \begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 + 3x_2 \\ b_1 &= \begin{bmatrix} 2 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1 - 4x_2 \\ b_1 &= \begin{bmatrix} 9 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 9x_1 - x_2 \end{aligned}$$

Hence,

$$\begin{bmatrix} 1 & 3 \\ 2 & -4 \\ 9 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 \\ 2x_1 - 4x_2 \\ 9x_1 - x_2 \end{bmatrix}$$

EXAMPLE 5

Let $A = \begin{bmatrix} 3 & 4 & -5 \\ 1 & 0 & 2 \end{bmatrix}$. Calculate $A\vec{x}$ where:

$$\text{(a) } \vec{x} = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}, \qquad \text{(b) } \vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \text{(c) } \vec{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Solution: We have

$$A\vec{x} = \begin{bmatrix} 3 & 4 & -5 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} = \begin{bmatrix} 3(2) + 4(-1) + (-5)(6) \\ 1(2) + (0)(-1) + (2)(6) \end{bmatrix} = \begin{bmatrix} -28 \\ 14 \end{bmatrix}$$

$$A\vec{x} = \begin{bmatrix} 3 & 4 & -5 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3(1) + 4(0) + (-5)(0) \\ 1(1) + (0)(0) + (2)(0) \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$A\vec{x} = \begin{bmatrix} 3 & 4 & -5 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3(0) + 4(0) + (-5)(1) \\ 1(0) + (0)(0) + (2)(1) \end{bmatrix} = \begin{bmatrix} -5 \\ 2 \end{bmatrix}$$

Coefficients are the Columns of A

Observe that by using operations on vectors in $\mathbb{R}^n$ the system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \vdots \qquad\qquad \vdots \quad &= \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be written as

$$x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

To write the right-hand side in the form $A\vec{x}$, we make the following definition.

DEFINITION

Matrix-Vector Multiplication

Let A be an $m \times n$ matrix whose columns are denoted $\vec{a}_i$ for $1 \leq i \leq n$. Then, for any $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, we define

$$A\vec{x} = x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

REMARKS

1. As above, if A is an $m \times n$ matrix, then $A\vec{x}$ only makes sense if $\vec{x} \in \mathbb{R}^n$. Moreover, if $\vec{x} \in \mathbb{R}^n$, then $A\vec{x} \in \mathbb{R}^m$.
2. Observe that this definition shows that for any $\vec{x} \in \mathbb{R}^n$, $A\vec{x}$ is a linear combination of the columns of A.

EXAMPLE 6

Let $A = \begin{bmatrix} 3 & 4 & -5 \\ 1 & 0 & 2 \end{bmatrix}$. Calculate $A\vec{x}$ where:

$$\text{(a) } \vec{x} = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}, \qquad \text{(b) } \vec{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Solution: We have

$$A\begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} = 2\begin{bmatrix} 3 \\ 1 \end{bmatrix} + (-1)\begin{bmatrix} 4 \\ 0 \end{bmatrix} + 6\begin{bmatrix} -5 \\ 2 \end{bmatrix} = \begin{bmatrix} 2(3) + (-1)(4) + (6)(-5) \\ 2(1) + (-1)(0) + (6)(2) \end{bmatrix} = \begin{bmatrix} -28 \\ 14 \end{bmatrix}$$

$$A\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 0\begin{bmatrix} 3 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 4 \\ 0 \end{bmatrix} + 0\begin{bmatrix} -5 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$$

Two Important Facts:

1. Observe that if $\vec{x}, \vec{y} \in \mathbb{R}^n$, then

$$\vec{x}^T\vec{y} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = [x_1y_1 + \cdots + x_ny_n] = \vec{x} \cdot \vec{y}$$

2. In most cases, we will now represent a system of linear equations $\left[\begin{array}{c|c} A & \vec{b} \end{array}\right]$ as $A\vec{x} = \vec{b}$. This will allow us to use properties of matrix vector multiplication when dealing with systems of linear equations.

Matrix Multiplication

DEFINITION

Matrix Multiplication

Let A be an $m \times n$ matrix and let $B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_p \end{bmatrix}$ be an $n \times p$ matrix. Then we define AB to be the $m \times p$ matrix

$$AB = A\begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_p \end{bmatrix} = \begin{bmatrix} A\vec{b}_1 & \cdots & A\vec{b}_p \end{bmatrix}$$

We can compute the entries of AB easily using our first interpretation of matrix-vector multiplication. In particular, if $\vec{a}_i^T$ are the rows of A, then the ij-th entry of AB is the i-th entry of $A\vec{b}_j$. That is

$$(AB)_{ij} = \vec{a}_i \cdot \vec{b}_j = \vec{a}_i^T \vec{b}_j = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} (A)_{ik}(B)_{kj}$$

Observe that the number of columns of A must equal the number of rows of B for this to be defined.

EXAMPLE 7

$$\begin{bmatrix} 1 & 3 & 2 \\ -1 & 0 & -3 \end{bmatrix} \begin{bmatrix} -2 & 4 \\ -4 & 5 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1(-2)+3(-4)+2(0) & 1(4)+3(5)+2(0) \\ (-1)(-2)+0(-4)+(-3)(0) & (-1)(4)+0(5)+(-3)(0) \end{bmatrix}$$
$$= \begin{bmatrix} -14 & 19 \\ 2 & -4 \end{bmatrix}$$

$$\begin{bmatrix} -2 & 4 \\ -4 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ -1 & 0 & -3 \end{bmatrix} = \begin{bmatrix} (-2)(1)+4(-1) & (-2)(3)+4(0) & (-2)(2)+4(-3) \\ (-4)(1)+5(-1) & (-4)(3)+5(0) & (-4)(2)+5(-3) \\ 0(1)+0(-1) & 0(3)+0(0) & 0(2)+0(-3) \end{bmatrix}$$
$$= \begin{bmatrix} -6 & -6 & -16 \\ -9 & -12 & -23 \\ 0 & 0 & 0 \end{bmatrix}$$

EXERCISE 2

Calculate the following matrix products, or state why they don't exist.

(a) $\begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 0 & 1 \\ 0 & 1 & -2 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 3 & 0 & 1 \\ 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$

(d) $\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$

THEOREM 3

If A, B, and C are matrices of the correct size so that the required products are defined, and $t \in \mathbb{R}$, then:

(1) $A(B+C) = AB + AC$
(2) $t(AB) = (tA)B = A(tB)$
(3) $A(BC) = (AB)C$
(4) $(AB)^T = B^T A^T$

Proof: We prove (1) and (3) and leave (2) and (4) as exercises.

For (1) we have

$$\begin{aligned}
\big(A(B+C)\big)_{ij} &= \sum_{k=1}^{n} (A)_{ik}(B+C)_{kj} \\
&= \sum_{k=1}^{n} (A)_{ik}\left((B)_{kj} + (C)_{kj}\right) \\
&= \sum_{k=1}^{n} (A)_{ik}(B)_{kj} + \sum_{k=1}^{n} (A)_{ik}(C)_{kj} \\
&= \big(AB\big)_{ij} + \big(AC\big)_{ij} \\
&= \big(AB + AC\big)_{ij}
\end{aligned}$$

For (3) we have

$$\begin{aligned}
\big(A(BC)\big)_{ij} &= \sum_{k=1}^{n} (A)_{ik}(BC)_{kj} \\
&= \sum_{k=1}^{n} (A)_{ik}\left[\sum_{\ell=1}^{n} (B)_{k\ell}(C)_{\ell j}\right] \\
&= \sum_{k=1}^{n}\left[\sum_{\ell=1}^{n} (A)_{ik}(B)_{k\ell}(C)_{\ell j}\right] \\
&= \sum_{\ell=1}^{n}\sum_{k=1}^{n} (A)_{ik}(B)_{k\ell}(C)_{\ell j} \\
&= \sum_{\ell=1}^{n}\left[\sum_{k=1}^{n} (A)_{ik}(B)_{k\ell}\right](C)_{\ell j} \\
&= \sum_{\ell=1}^{n} (AB)_{i\ell}(C)_{\ell j} \\
&= \big((AB)C\big)_{ij}
\end{aligned}$$

$\square$

Observe that Example 7 shows that in general $AB \neq BA$. When multiplying an equation by a matrix we have to make sure we that we keep this in mind. That is, if $A = B$ and we multiply both sides by a matrix D, then we get either $DA = DB$ or $AD = BD$. In general $AD \neq DB$. Similarly, we do not have the cancellation law for matrix multiplication; if $AC = BC$, then we can not guarantee that $A = B$. This is demonstrated in the next example.

EXAMPLE 8

We have

$$\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 15 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

but

$$\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 15 \\ 0 & -5 \end{bmatrix}$$

However, in many cases we will like to prove that two matrices are equal. To do this, we often use the following theorem.

THEOREM 4

Suppose that A and B are $m \times n$ matrices such that $A\vec{x} = B\vec{x}$ for every $\vec{x} \in \mathbb{R}^n$. Then $A = B$.

Proof: Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_n \end{bmatrix}$. Let $\vec{e}_i$ denote the i-th standard basis vector. Then, since $A\vec{x} = B\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$, we have that

$$\vec{a}_i = A\vec{e}_i = B\vec{e}_i = \vec{b}_i$$

for $1 \leq i \leq n$. Hence $A = B$. $\square$

Identity Matrix

When considering multiplication, we are often interested in finding the multiplicative identity; the element I such that $A \times I = A = I \times A$ for every element A. For $n \times n$ matrices, the multiplicative identity is the $n \times n$ **identity matrix**.

DEFINITION
Identity Matrix

The $n \times n$ **identity matrix**, denoted by I or I_n, is the matrix such that $(I)_{ii} = 1$ for $1 \leq i \leq n$ and $(I)_{ij} = 0$ whenever $i \neq j$. In particular, the columns of I_n are the standard basis vectors of $\mathbb{R}^n$.

EXAMPLE 9

For any 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we have

$$AI = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

REMARK

Observe that the set $M_{m\times n}(\mathbb{R})$ with $m \neq n$ does not have a multiplicative identity since if $AB = A$, then B must be an $n\times n$ matrix, while if $CA = A$, then C must be an $m \times m$ matrix.

THEOREM 5

Let A be an $m \times n$ matrix. Prove that $I_m A = A$ and $AI_n = A$.

Proof: The proof is left as an exercise.

This theorem shows that I_n is the multiplicative identity of $M_{n\times n}(\mathbb{R})$.

Block Multiplication

So far in this section, we have often made use of the fact that we could represent a matrix as a list of columns ($n \times 1$ matrices) or as a list of rows ($1 \times n$ matrices). Sometimes it can also be very useful to represent a matrix as a group of submatrices called **blocks**.

DEFINITION
Block Matrix

If A is an $m \times n$ matrix, then we can write A as the $k \times \ell$ **block matrix**

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1\ell} \\ \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{k\ell} \end{bmatrix}$$

where A_{ij} is a block such that all blocks in the i-th row have the same number of rows and all blocks in the j-th column have the same number of columns.

EXAMPLE 10

Let $A = \begin{bmatrix} 1 & -1 & 3 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & 2 \end{bmatrix}$. Then, we can write A as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where $A_{11} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 3 & 4 \\ 0 & 0 \end{bmatrix}$, $A_{21} = \begin{bmatrix} 0 & 3 \end{bmatrix}$, and $A_{22} = \begin{bmatrix} 1 & 2 \end{bmatrix}$. Or, we could write A as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}$$

where $A_{11} = \begin{bmatrix} 1 & -1 & 3 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 4 \end{bmatrix}$, $A_{21} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$, $A_{22} = \begin{bmatrix} 0 \end{bmatrix}$, $A_{31} = \begin{bmatrix} 0 & 3 & 1 \end{bmatrix}$, and $A_{32} = \begin{bmatrix} 2 \end{bmatrix}$.

Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. If we write A as a $k \times \ell$ block matrix and B as an $\ell \times s$ block matrix, then we can calculate AB, by multiplying the block matrices together using the definition of matrix multiplication. For example, if $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}$ and $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$, then

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \\ A_{31}B_{11} + A_{32}B_{21} & A_{31}B_{12} + A_{32}B_{22} \end{bmatrix}$$

assuming that we have divided A and B into blocks such that all of the required products of the blocks is defined.

EXAMPLE 11

Let $A = \begin{bmatrix} 1 & 2 & -3 \\ 0 & 3 & 1 \\ 1 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 3 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 3 \end{bmatrix}$. Let $A_{11} = \begin{bmatrix} 1 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 2 & -3 \end{bmatrix}$, $A_{21} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $A_{22} = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ so that $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, and let $B_{11} = \begin{bmatrix} 2 \end{bmatrix}$, $B_{12} = \begin{bmatrix} 3 & 1 \end{bmatrix}$, $B_{21} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and $B_{22} = \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix}$, so that $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$. Then we have

$$\begin{aligned} AB &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} \end{aligned}$$

Computing each entry we get

$$\begin{aligned} A_{11}B_{11} + A_{12}B_{21} &= \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 2 \end{bmatrix} + \begin{bmatrix} 2 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix} = \begin{bmatrix} 2 \end{bmatrix} \\ A_{11}B_{12} + A_{12}B_{22} &= \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -3 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -13 \end{bmatrix} = \begin{bmatrix} 5 & -12 \end{bmatrix} \\ A_{21}B_{11} + A_{22}B_{21} &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \end{bmatrix} + \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \\ A_{21}B_{12} + A_{22}B_{22} &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 3 & -3 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 3 & -3 \\ 3 & 7 \end{bmatrix} \end{aligned}$$

Hence,

$$AB = \begin{bmatrix} 2 & 5 & -12 \\ 0 & 3 & -3 \\ 2 & 3 & 7 \end{bmatrix}$$

REMARKS

1. If you multiply out AB like normal and compare this to the calculations above it becomes clear why block multiplication works.
2. Block multiplication is used to distribute matrix multiplication over multiple computers, and in a variety of applications.

3.2 Linear Mappings

Recall that a function $f : A \to B$ is a rule that associates with each element $a \in A$ one element $f(a) \in B$ called the **image** of a under f. The subset of A for which $f(a)$ is defined is called the **domain** of f. The set B is called the **codomain** of f while the subset of B consisting of all $f(a)$ is called the **range** of f and is denoted $R(f)$.

Matrix Mappings

Observe that if A is an $m \times n$ matrix, then we can define a function $f : \mathbb{R}^n \to \mathbb{R}^m$ by $f(\vec{x}) = A\vec{x}$ called a **matrix mapping**.

REMARK

As we indicated in the previous section, if $f : \mathbb{R}^n \to \mathbb{R}^m$, then we should write

$$f\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

However, we often write $f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)$ as is typically done in other areas.

EXAMPLE 1

Let $A = \begin{bmatrix} 1 & 3 & 2 \\ -1 & 3 & 1 \end{bmatrix}$ and define $f(\vec{x}) = A\vec{x}$. Observe that for $A\vec{x}$ to be defined, $\vec{x}$ must be a vector in $\mathbb{R}^3$. Moreover, $A\vec{x}$ is defined for every $\vec{x} \in \mathbb{R}^3$, so the domain of f is $\mathbb{R}^3$. Also, we see that $A\vec{x} \in \mathbb{R}^2$ and so the codomain of f is $\mathbb{R}^2$. We can easily compute the images of some vectors in $\mathbb{R}^3$ under f using either definition of matrix vector multiplication.

$$f(4,-2,5) = \begin{bmatrix} 1 & 3 & 2 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 5 \end{bmatrix} = (8,-5)$$

$$f(-1,-1,2) = \begin{bmatrix} 1 & 3 & 2 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} = (0,0)$$

$$f(1,0,0) = \begin{bmatrix} 1 & 3 & 2 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = (1,-1)$$

$$f(x_1,x_2,x_3) = \begin{bmatrix} 1 & 3 & 2 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (x_1+3x_2+2x_3, -x_1+3x_2+x_3)$$

Using the properties of matrix multiplication, we prove the following important properties of matrix mappings.

THEOREM 1

Let A be an $m \times n$ matrix, and let $f : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $f(\vec{x}) = A\vec{x}$. Then for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $b, c \in \mathbb{R}$ we have

$$f(b\vec{x} + c\vec{y}) = bf(\vec{x}) + cf(\vec{y})$$

Proof: Using Theorem 3.1.1 we get

$$f(b\vec{x} + c\vec{y}) = A(b\vec{x} + c\vec{y}) = A(b\vec{x}) + A(c\vec{y}) = bA\vec{x} + cA\vec{y} = bf(\vec{x}) + cf(\vec{y})$$

$\square$

Any function which has the property that $f(b\vec{x} + c\vec{y}) = bf(\vec{x}) + cf(\vec{y})$ is said to **preserve linear combinations**. This property, called the **linearity property**, is extremely important and useful in Linear Algebra. We now look at functions which have this property.

Linear Mappings

DEFINITION

Linear Mapping

A function $L : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a **linear mapping** if for every $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $b, c \in \mathbb{R}$ we have

$$L(b\vec{x} + c\vec{y}) = bL(\vec{x}) + cL(\vec{y})$$

REMARKS

1. A linear mapping may also be called a linear transformation.
2. A linear mapping $L : \mathbb{R}^n \to \mathbb{R}^n$ is sometimes called a **linear operator**. This often done when we wish to stress the fact that the linear mapping has the same domain and codomain.

Observe that if we have a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ and that $\vec{x}$ can be written as a linear combination of vectors $\vec{v}_1, \ldots, \vec{v}_k$, say $\vec{x} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$, then the image of $\vec{x}$ under L can be written as a linear combination of the images of the vectors $\vec{v}_1, \ldots, \vec{v}_k$ under L. That is

$$L(\vec{x}) = L(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) = c_1L(\vec{v}_1) + \cdots + c_kL(\vec{v}_k)$$

since L preserves linear combinations. This is the most important property of linear mappings; it will be used frequently!

EXAMPLE 2 Prove that the function $L : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $L(x_1, x_2, x_3) = (3x_1 - x_2, 2x_1 + 2x_3)$ is a linear mapping.

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

$$\begin{aligned} L(b\vec{x} + c\vec{y}) &= L\big(b(x_1, x_2, x_3) + c(y_1, y_2, y_3)\big) \\ &= L(bx_1 + cy_1, bx_2 + cy_2, bx_3 + cy_3) \\ &= \big(3(bx_1 + cy_1) - (bx_2 + cy_2), 2(bx_1 + cy_1) + 2(bx_3 + cy_3)\big) \\ &= b(3x_1 - x_2, 2x_1 + 2x_3) + c(3y_1 - y_2, 2y_1 + 2y_3) \\ &= bL(\vec{x}) + cL(\vec{y}) \end{aligned}$$

EXAMPLE 3 Prove that $L : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $L(x_1, x_2) = (x_1^2 - x_2^2, x_1x_2)$ is not linear.

Solution: To prove L is not linear we just need to find one example to show that L does not preserve linear combinations. Observe that $L(1, 2) = (-3, 2)$ and $L(2, 4) = (-12, 8)$, but $2L(1, 2) = (-6, 4) \neq L\big(2(1, 2)\big) = L(2, 4)$. Hence, L is not linear.

EXAMPLE 4 Let $\vec{a} \in \mathbb{R}^n$. Show that $\operatorname{proj}_{\vec{a}} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear mapping.

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

$$\begin{aligned} \operatorname{proj}_{\vec{a}}(b\vec{x} + c\vec{y}) &= \frac{(b\vec{x} + c\vec{y}) \cdot \vec{a}}{\|\vec{a}\|^2}\vec{a} \\ &= \frac{b(\vec{x} \cdot \vec{a})}{\|\vec{a}\|^2}\vec{a} + \frac{c(\vec{y} \cdot \vec{a})}{\|\vec{a}\|^2}\vec{a} \\ &= b\operatorname{proj}_{\vec{a}}\vec{x} + c\operatorname{proj}_{\vec{a}}\vec{y} \end{aligned}$$

Hence, $\operatorname{proj}_{\vec{a}}$ is linear.

EXERCISE 1

Prove that the mapping $L : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $L(x_1, x_2, x_3) = (x_1 - x_2, x_1 - 2x_3)$ is linear.

Theorem 3.1.1 shows that every matrix mapping is a linear mapping. It is natural to ask whether the converse is true: "Is every linear mapping a matrix mapping?" To answer this question, we will see that our second interpretation of matrix vector multiplication is very important.

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping and let $\vec{x} \in \mathbb{R}^n$. We know that $\vec{x}$ can be written as a unique linear combination of the standard basis vectors, say

$$\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$$

Hence, using the linearity property we get

$$L(\vec{x}) = L(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) = x_1L(\vec{e}_1) + \cdots + x_nL(\vec{e}_n)$$

Our second interpretation of matrix vector multiplication says that a matrix times a vector gives us a linear combination of the columns of the matrix. Using this in reverse, we see that

$$L(\vec{x}) = \begin{bmatrix} L(\vec{e}_1) & \cdots & L(\vec{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

We have proven the following theorem.

THEOREM 2

Every linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ can be represented as a matrix mapping with matrix whose columns are the images of the standard basis vectors of $\mathbb{R}^n$ under L. That is, $L(\vec{x}) = [L]\vec{x}$ where

$$[L] = \begin{bmatrix} L(\vec{e}_1) & \cdots & L(\vec{e}_n) \end{bmatrix}$$

DEFINITION
Standard Matrix

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then the matrix $[L] = \begin{bmatrix} L(\vec{e}_1) & \cdots & L(\vec{e}_n) \end{bmatrix}$ is called the **standard matrix** of L.

REMARK

We are calling this the standard matrix of L since it is based on the standard basis. In Math 235, we will see that we can find the matrix representation of a linear mapping with respect to different bases.

EXAMPLE 5

Determine the standard matrix of $L(x_1, x_2, x_3) = (3x_1 - x_2, 2x_1 + 2x_3)$.

Solution: By definition, the columns of the standard matrix $[L]$ are the images of the standard basis vectors of $\mathbb{R}^3$ under L. We have

$$\begin{aligned} L(1,0,0) &= (3,2) \\ L(0,1,0) &= (-1,0) \\ L(0,0,1) &= (0,2) \end{aligned}$$

Thus,

$$[L] = \begin{bmatrix} L(\vec{e}_1) & L(\vec{e}_2) & L(\vec{e}_3) \end{bmatrix} = \begin{bmatrix} 3 & -1 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

REMARK

Be careful to remember that when we write $L(1,0,0) = (3,2)$, we really mean $L(\vec{e}_1) = L\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ so that $L(\vec{e}_1)$ is in fact a column of $[L]$.

EXAMPLE 6

Let $\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Determine the standard matrix of $\operatorname{proj}_{\vec{a}} : \mathbb{R}^2 \to \mathbb{R}^2$.

Solution: We need to find the projection of the standard basis vectors of $\mathbb{R}^2$ onto $\vec{a}$.

$$\begin{aligned} \operatorname{proj}_{\vec{a}}(\vec{e}_1) &= \frac{\vec{e}_1 \cdot \vec{a}}{\|\vec{a}\|^2}\vec{a} = \frac{1}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 2/5 \end{bmatrix} \\ \operatorname{proj}_{\vec{a}}(\vec{e}_2) &= \frac{\vec{e}_2 \cdot \vec{a}}{\|\vec{a}\|^2}\vec{a} = \frac{2}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2/5 \\ 4/5 \end{bmatrix} \end{aligned}$$

Hence, $[\operatorname{proj}_{\vec{a}}] = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}$.

EXERCISE 2

Determine the standard matrix of the linear mapping $L(x_1, x_2, x_3) = (x_1 - x_2, x_1 - 2x_3)$.

We now look at two important examples of linear mappings which will be used throughout Math 136 and Math 235.

Rotations in $\mathbb{R}^2$

Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ denote the function that rotates a vector $\vec{x} \in \mathbb{R}^2$ about the origin through an angle θ. Using basic trigonometric identities, it is easy to show that

$$R_\theta(x_1, x_2) = (x_1 \cos\theta - x_2 \sin\theta, x_1 \sin\theta + x_2 \cos\theta)$$

and that R_θ is linear. We then find that the standard matrix of R_θ is

$$[R_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

EXAMPLE 7 Determine the standard matrix of $R_{\pi/4}$.

Solution: We have

$$[R_{\pi/4}] = \begin{bmatrix} \cos\pi/4 & -\sin\pi/4 \\ \sin\pi/4 & \cos\pi/4 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

We call R_θ a **rotation** in $\mathbb{R}^2$ and we call its standard matrix $[R_\theta]$ a **rotation matrix**. The following theorem shows some important properties of a rotation matrix.

THEOREM 3 Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation with rotation matrix $A = [R_\theta]$. Then the columns of A are orthogonal unit vectors.

Proof: The columns of A are $\vec{a}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and $\vec{a}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$. Hence,

$$\begin{aligned} \vec{a}_1 \cdot \vec{a}_2 &= -\cos\theta\sin\theta + \cos\theta\sin\theta = 0 \\ \|\vec{a}_1\|^2 &= \cos^2\theta + \sin^2\theta = 1 \\ \|\vec{a}_2\|^2 &= \cos^2\theta + \sin^2\theta = 1 \end{aligned}$$

$\square$

Reflections

Let $\operatorname{refl}_{\vec{n}} : \mathbb{R}^n \to \mathbb{R}^n$ denote the linear mapping which maps a vector $\vec{x}$ to its mirror image in the hyperplane with normal vector $\vec{n}$. Figure 3.2.1 below shows a reflection over a line in $\mathbb{R}^2$. From the figure, we see that the reflection of $\vec{x}$ over the line with normal vector $\vec{n}$ is given by

$$\operatorname{refl}_{\vec{n}} \vec{x} = \vec{x} - 2\operatorname{proj}_{\vec{n}} \vec{x}$$

We extend this formula directly to the n-dimensional case.

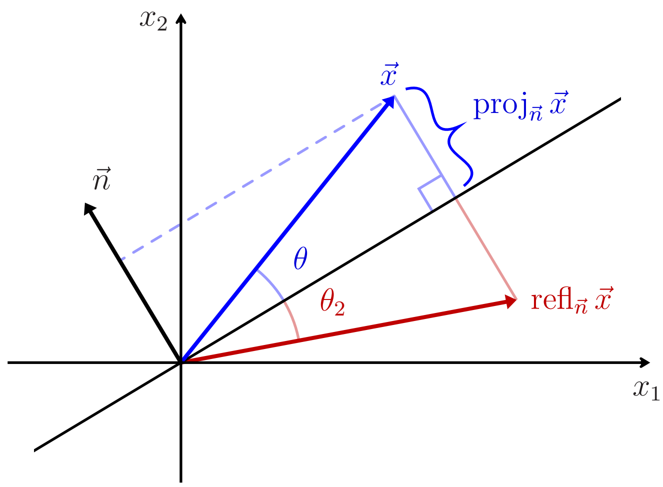

Figure 3.2.1: Reflection of $\vec{x}$ across the line with normal vector $\vec{n}$

EXAMPLE 8

Determine the reflection of $\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 3 \end{bmatrix}$ over the hyperplane in $\mathbb{R}^4$ with normal vector $\vec{n} = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 0 \end{bmatrix}$.

Solution: We have

$$\begin{aligned} \text{refl}_{\vec{n}}\,\vec{x} &= \vec{x} - 2\,\text{proj}_{\vec{n}}\,\vec{x} = \vec{x} - 2\frac{\vec{x}\cdot\vec{n}}{\|\vec{n}\|^2}\vec{n} \\ &= \begin{bmatrix} 1 \\ -1 \\ 1 \\ 3 \end{bmatrix} - \frac{-6}{9}\begin{bmatrix} 1 \\ 2 \\ -2 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 5/3 \\ 1/3 \\ -1/3 \\ 3 \end{bmatrix} \end{aligned}$$

EXAMPLE 9

Determine the standard matrix of $\text{refl}_{\vec{n}} : \mathbb{R}^3 \to \mathbb{R}^3$ where $\vec{n} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$.

Solution: We have

$$\begin{aligned}
\operatorname{refl}_{\vec{n}} \vec{e}_1 &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{2}{14} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 6/7 \\ -2/7 \\ 3/7 \end{bmatrix} \\
\operatorname{refl}_{\vec{n}} \vec{e}_2 &= \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{4}{14} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -2/7 \\ 3/7 \\ 6/7 \end{bmatrix} \\
\operatorname{refl}_{\vec{n}} \vec{e}_3 &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \frac{6}{14} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 3/7 \\ 6/7 \\ -2/7 \end{bmatrix}
\end{aligned}$$

Hence,

$$[\operatorname{refl}_{\vec{n}}] = \begin{bmatrix} 6/7 & -2/7 & 3/7 \\ -2/7 & 3/7 & 6/7 \\ 3/7 & 6/7 & -2/7 \end{bmatrix}$$

3.3 Special Subspaces

When dealing with functions, it is often important and useful to look at their domain and range. Thus, we now look at the domain and range of linear mappings.

Kernel

We have seen that unlike many functions you deal with in Calculus, linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$ never have restricted domain. That is, the domain of L is always all of $\mathbb{R}^n$. However, there is an important subset of the domain which we will use frequently.

DEFINITION
Kernel

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then the **kernel** of L is defined to be

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

EXAMPLE 1

Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $L(x_1, x_2) = (2x_1 - 2x_2, -x_1 + x_2)$. Determine which of the following vectors is in the kernel of L: $\vec{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\vec{x}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $\vec{x}_3 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

Solution: We have

$$\begin{aligned}
L(\vec{x}_1) &= L(0,0) = (0,0) \\
L(\vec{x}_2) &= L(2,2) = (0,0) \\
L(\vec{x}_3) &= L(3,-1) = (8,-4)
\end{aligned}$$

Hence, $\vec{x}_1, \vec{x}_2 \in \ker(L)$ while $\vec{x}_3 \notin \ker(L)$.

EXAMPLE 2 Determine the kernel of $L(x_1, x_2, x_3) = (3x_1 - x_2, 2x_1 + 2x_3)$.

Solution: By definition, the kernel of L is the set of all $\vec{x} \in \mathbb{R}^3$ such that $L(\vec{x}) = \vec{0}$. That is, $\vec{x} \in \ker(L)$ if

$$(0, 0) = L(x_1, x_2, x_3) = (3x_1 - x_2, 2x_1 + 2x_3)$$

Hence, every $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in the kernel of L must satisfy

$$\begin{aligned} 3x_1 - x_2 &= 0 \\ 2x_1 + 2x_3 &= 0 \end{aligned}$$

We find the general solution of the homogeneous system is

$$\vec{x} = t \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix} = \operatorname{Span} \left\{ \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix} \right\}$$

Thus,

$$\ker(L) = \operatorname{Span} \left\{ \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix} \right\}$$

EXAMPLE 3 Determine the kernel of $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$.

Solution: We need to find all $\vec{x} \in \mathbb{R}^2$ such that $R_\theta(\vec{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Recall that $[R_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Thus, we need to solve

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = R_\theta(\vec{x}) = [R_\theta]\vec{x} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \vec{x}$$

This is a homogeneous system, so we row reduce the coefficient matrix.

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Hence, the system has the unique solution $\vec{x} = \vec{0}$. So,

$$\ker(R_\theta) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$$

Note that geometrically this result is obvious.

EXERCISE 1 Determine the kernel of $L(x_1, x_2, x_3) = (x_1 - x_2, x_1 - 2x_3)$.

Observe that in both of these examples the kernel of the linear mapping is a subspace of the appropriate $\mathbb{R}^n$. To prove that this is always the case, we will use the following lemma.

LEMMA 1 Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then $L(\vec{0}) = \vec{0}$.

Proof: The proof is left as an exercise.

THEOREM 2 Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then $\ker(L)$ is a subspace of $\mathbb{R}^n$.

Proof: We show that $\ker(L)$ satisfies the definition of a subspace of $\mathbb{R}^n$. We first observe that by definition $\ker(L)$ is a subset of $\mathbb{R}^n$. Also, we have that $\vec{0} \in \ker(L)$ by Lemma 1. Let $\vec{x}, \vec{y} \in \ker(L)$. Then $L(\vec{x}) = \vec{0}$ and $L(\vec{y}) = \vec{0}$ so

$$L(\vec{x} + \vec{y}) = L(\vec{x}) + L(\vec{y}) = \vec{0} + \vec{0} = \vec{0}$$

so $\vec{x} + \vec{y} \in \ker(L)$. Thus, $\ker(L)$ is closed under addition. Similarly, for any $c \in \mathbb{R}$ we have

$$L(c\vec{x}) = cL(\vec{x}) = c\vec{0} = \vec{0}$$

so $c\vec{x} \in \ker(L)$. Therefore, $\ker(L)$ is also closed under scalar multiplication. Hence, we have shown that $\ker(L)$ is a subspace of $\mathbb{R}^n$. $\square$

Range

DEFINITION
Range

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. The **range** of L is

$$R(L) = \{L(\vec{x}) \in \mathbb{R}^m \mid \vec{x} \in \mathbb{R}^n\}$$

EXAMPLE 4 Let $L : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $L(x_1, x_2) = (x_1 + x_2, x_1 - x_2, x_2)$. Determine which of the following vectors is in the range of L: $\vec{y}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, $\vec{y}_2 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$, $\vec{y}_3 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$.

Solution: By Lemma 1, we have $L(0, 0) = (0, 0, 0)$, so $\vec{y}_1 \in R(L)$.

For $\vec{y}_2$, we need to find if there exist x_1 and x_2 such that

$$(2, 3, 1) = L(x_1, x_2) = (x_1 + x_2, x_1 - x_2, x_2)$$

That is, we need to solve the system of equations

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

This system is inconsistent so $\vec{y}_2 \notin R(L)$.

For $\vec{y}_3$, we need to find if there exist x_1 and x_2 such that

$$(3, -1, 2) = L(x_1, x_2) = (x_1 + x_2, x_1 - x_2, x_2)$$

That is, we need to solve the system of equations

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

We find that the system is consistent with unique solution is $x_1 = 1$, $x_2 = 2$. Hence, $L(1, 2) = (3, -1, 2)$, so $\vec{y}_3 \in R(L)$.

EXAMPLE 5

Let $\vec{a} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Determine the range of $\operatorname{proj}_{\vec{a}} : \mathbb{R}^2 \to \mathbb{R}^2$.

Solution: We see that every vector $\operatorname{proj}_{\vec{a}} \vec{x}$ in the range of $\operatorname{proj}_{\vec{a}}$ has the form

$$\operatorname{proj}_{\vec{a}} \vec{x} = \frac{\vec{x} \cdot \vec{a}}{\|\vec{a}\|^2} \vec{a} = c\vec{a}$$

for some $c \in \mathbb{R}$. Moreover, for any $c \in \mathbb{R}$, we can find some $\vec{x} \in \mathbb{R}^2$ such that $c = \frac{\vec{x} \cdot \vec{a}}{\|\vec{a}\|^2}$. Thus,

$$R(\operatorname{proj}_{\vec{a}}) = \operatorname{Span}\{\vec{a}\}$$

EXAMPLE 6

Determine the range of $L(x_1, x_2, x_3) = (3x_1 - x_2, 2x_1 + 2x_3)$.

Solution: We see that every vector $L(\vec{x})$ has the form

$$\begin{bmatrix} 3x_1 - x_2 \\ 2x_1 + 2x_3 \end{bmatrix} = x_1 \begin{bmatrix} 3 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

Hence,

$$R(L) = \operatorname{Span}\left\{ \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$$

Observe that the spanning set for $R(L)$ is linearly dependent and so we could simplify this to be, for example,

$$R(L) = \operatorname{Span}\left\{ \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$$

EXERCISE 2 Determine the range of $L(x_1, x_2, x_3) = (x_1 - x_2, x_1 - 2x_3)$.

Observe in the examples above that the range of a linear mapping is a subspace of the codomain. Of course, this is true for all linear mappings.

THEOREM 3 Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Then, $R(L)$ is a subspace of $\mathbb{R}^m$.

Proof: The proof is left as an exercise.

In many cases we are interested in whether the range of a linear mapping L equals its codomain. That is, for every vector $\vec{y}$ in the codomain, can we find a vector $\vec{x}$ in the domain such that $L(\vec{x}) = \vec{y}$.

REMARK

A function whose range equals its codomain is said to be **onto**. We will look more closely at onto mappings in Math 235.

EXAMPLE 7 Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $L(x_1, x_2, x_3) = (x_1 + x_2 + x_3, 2x_1 - 2x_3, x_2 + 3x_3)$. Prove that $R(L) = \mathbb{R}^3$.

Solution: To prove that $R(L) = \mathbb{R}^3$, we need to show that if we pick any vector $\vec{y} = \mathbb{R}^3$, then we can find a vector $\vec{x} \in \mathbb{R}^3$ such that $L(\vec{x}) = \vec{y}$. Let $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \in \mathbb{R}^3$. Consider

$$(y_1, y_2, y_3) = L(x_1, x_2, x_3) = (x_1 + x_2 + x_3, 2x_1 - 2x_3, x_2 + 3x_3)$$

This gives the system of linear equations

$$\begin{aligned} x_1 + x_2 + x_3 &= y_1 \\ 2x_1 - 2x_3 &= y_2 \\ x_2 + 3x_3 &= y_3 \end{aligned}$$

Row reducing the corresponding augmented matrix gives

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & y_1 \\ 2 & 0 & -2 & y_2 \\ 0 & 1 & 3 & y_3 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 1 & 1 & y_1 \\ 0 & -2 & -3 & -2y_1 + y_2 \\ 0 & 0 & 1 & -y_1 + \frac{1}{2}y_2 + y_3 \end{array}\right]$$

Therefore $R(L) = \mathbb{R}^3$ since the system is consistent for all y_1, y_2, and y_3. In particular, solving the system we find that by taking $x_1 = -y_1 + y_2 + y_3$, $x_2 = 3y_1 - \frac{3}{2}y_2 - 2y_3$, $x_3 = -y_1 + \frac{1}{2}y_2 + y_3$ we get

$$L(x_1, x_2, x_3) = (y_1, y_2, y_3)$$

EXAMPLE 8

Let A be an $n \times n$ matrix and let $L : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $L(\vec{x}) = A\vec{x}$. Prove that $R(L) = \mathbb{R}^n$ if and only if $\operatorname{rank} A = n$.

Solution: Assume that $R(L) = \mathbb{R}^n$. Then, for every $\vec{y} \in \mathbb{R}^n$ there exists a vector $\vec{x}$ such that $\vec{y} = L(\vec{x}) = A\vec{x}$. Hence, the system of equations $A\vec{x} = \vec{y}$ with coefficient matrix A is consistent for every $\vec{y} \in \mathbb{R}^n$. By Theorem 2.2.3 this implies that $\operatorname{rank}(A) = n$.

On the other hand, if $\operatorname{rank} A = n$, then $L(\vec{x}) = A\vec{x} = \vec{y}$ is consistent for every $\vec{y} \in \mathbb{R}^n$, so $R(L) = \mathbb{R}^n$.

The Four Fundamental Subspaces of a Matrix

We now look at the relationship of the standard matrix of a linear mapping to its range and kernel. In doing so, we will derive the **four fundamental subspaces** of a matrix.

THEOREM 4

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping and let $A = [L]$ be the standard matrix of L. Then, $\vec{x} \in \ker(L)$ if and only if $A\vec{x} = \vec{0}$.

Proof: If $\vec{x} \in \ker(L)$, then $\vec{0} = L(\vec{x}) = A\vec{x}$. On the other hand, if $A\vec{x} = \vec{0}$, then $\vec{0} = A\vec{x} = L(\vec{x})$, so $\vec{x} \in \ker(L)$. $\square$

Since $\ker(L)$ is a subspace of $\mathbb{R}^n$, this theorem shows that for any $m \times n$ matrix A the set of all vectors $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{0}$ is a subspace of $\mathbb{R}^n$. This motivates the following definition.

DEFINITION
Nullspace

Let A be an $m \times n$ matrix. The set of all $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{0}$ is called the **nullspace** of A. We write

$$\operatorname{Null}(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}$$

EXAMPLE 9

Let $A = \begin{bmatrix} 1 & 3 & 1 \\ -1 & -2 & 2 \end{bmatrix}$. Determine a spanning set for the nullspace of A.

Solution: We need to find all $\vec{x}$ such that $A\vec{x} = \vec{0}$. We observe that this is a homogeneous system with coefficient matrix A. Row reducing gives

$$\begin{bmatrix} 1 & 3 & 1 \\ -1 & -2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -8 \\ 0 & 1 & 3 \end{bmatrix}$$

Thus, the general solution to the system is

$$\vec{x} = t \begin{bmatrix} 8 \\ -3 \\ 1 \end{bmatrix}$$

Hence,

$$\text{Null}(A) = \text{Span}\left\{ \begin{bmatrix} 8 \\ -3 \\ 1 \end{bmatrix} \right\}$$

EXAMPLE 10

Let $A = \begin{bmatrix} 1 & 2 & -1 \\ 5 & 0 & -4 \\ -2 & 6 & 4 \end{bmatrix}$. Show that $\text{Null}(A) = \{\vec{0}\}$.

Solution: We need to show that the only solution to the homogeneous system $A\vec{x} = \vec{0}$ is $\vec{x} = \vec{0}$. Row reducing the associated coefficient matrix we get

$$\begin{bmatrix} 1 & 2 & -1 \\ 5 & 0 & -4 \\ -2 & 6 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 \\ 0 & -10 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

Hence, the rank of the coefficient matrix equals the number of columns so by Theorem 2.2.3 there is a unique solution. Therefore, since the system is homogeneous, the only solution is $\vec{x} = \vec{0}$. Thus, $\text{Null}(A) = \{\vec{0}\}$.

THEOREM 5

Let A be an $m \times n$ matrix. A consistent system of linear equations $A\vec{x} = \vec{b}$ has a unique solution if and only if $\text{Null}(A) = \{\vec{0}\}$.

Proof: If $A\vec{x} = \vec{b}$ has a unique solution, then by Theorem 2.2.3, $\text{rank}\, A = n$. Hence, $A\vec{x} = \vec{0}$ has a unique solution, so $\text{Null}(A) = \{\vec{0}\}$.

If $\text{Null}(A) = \{\vec{0}\}$, then $A\vec{x} = \vec{0}$ has a unique solution, so by Theorem 2.2.3, $\text{rank}\, A = n$. Thus, $A\vec{x} = \vec{b}$ has a unique solution. $\square$

The connection between the range of a linear mapping and its standard matrix follows easily from our second interpretation of matrix vector multiplication.

THEOREM 6

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping with standard matrix $[L] = A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$. Then,

$$R(L) = \text{Span}\{\vec{a}_1, \ldots, \vec{a}_n\}$$

Proof: We have

$$L(\vec{x}) = A\vec{x} = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n$$

for all $x_1, \ldots, x_n \in \mathbb{R}^n$, and the result follows. $\square$

Therefore, the range of a linear mapping is the subspace spanned by the columns of its standard matrix.

DEFINITION
Columnspace

Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix}$. The **columnspace** of A is the subspace of $\mathbb{R}^m$ defined by

$$\operatorname{Col}(A) = \operatorname{Span}\{\vec{a}_1, \ldots, \vec{a}_n\} = \{A\vec{x} \in \mathbb{R}^m \mid \vec{x} \in \mathbb{R}^n\}$$

EXAMPLE 11

Let $A = \begin{bmatrix} 1 & 5 & -3 \\ 9 & 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -3 \\ 0 & 2 \\ 0 & 1 \end{bmatrix}$, then $\operatorname{Col}(A) = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 9 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \end{bmatrix} \right\}$ and $\operatorname{Col}(B) = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} \right\}$.

THEOREM 7

Let A be an $m \times n$ matrix. Then $\operatorname{Col}(A) = \mathbb{R}^m$ if and only if $\operatorname{rank}(A) = m$.

Proof: The proof is left as an exercise.

Recall that the transpose of a matrix changes the columns of a matrix into rows and vice versa. Hence, the subspace of $\mathbb{R}^n$ spanned by the rows of an $m \times n$ matrix A is just the columnspace of A^T.

DEFINITION
Rowspace

Let A be an $m \times n$ matrix. The **rowspace** of A is the subspace of $\mathbb{R}^n$ defined by

$$\operatorname{Row}(A) = \{A^T\vec{x} \in \mathbb{R}^n \mid \vec{x} \in \mathbb{R}^m\}$$

EXAMPLE 12

Let $A = \begin{bmatrix} 1 & 5 & -3 \\ 9 & 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -3 \\ 0 & 2 \\ 0 & 1 \end{bmatrix}$, then $\operatorname{Row}(A) = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}, \begin{bmatrix} 9 \\ 2 \\ 1 \end{bmatrix} \right\}$ and $\operatorname{Row}(B) = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$.

Lastly, since we have looked at the columnspace of A^T it makes sense to also consider the nullspace of A^T.

DEFINITION
Left Nullspace

Let A be an $m \times n$ matrix. The **left nullspace** of A is the subspace of $\mathbb{R}^m$ defined by

$$\operatorname{Null}(A^T) = \{\vec{x} \in \mathbb{R}^m \mid A^T\vec{x} = \vec{0}\}$$

EXAMPLE 13

Let $A = \begin{bmatrix} 1 & -3 \\ 0 & 2 \\ 0 & 1 \end{bmatrix}$. Then, $A^T = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 2 & 1 \end{bmatrix}$. We find that the general solution to the homogeneous system $A^T\vec{x} = \vec{0}$ is

$$\vec{x} = t\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}$$

Therefore,

$$\operatorname{Null}(A^T) = \operatorname{Span}\left\{\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}\right\}$$

For any $m \times n$ matrix A, we call the nullspace, columnspace, rowspace, and left nullspace the **four fundamental subspaces** of A. The next two theorems give you a brief glimpse of why these subspaces are useful. This will be explored further in Math 235

EXAMPLE 14

Find the four fundamental subspaces of $A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 5 & 2 \end{bmatrix}$.

Solution: To find $\operatorname{Null}(A)$ we solve $A\vec{x} = \vec{0}$. Row reducing A gives

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

Hence, $\operatorname{Null}(A) = \operatorname{Span}\left\{\begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}\right\}$.

For the columnspace of A we have

$$\operatorname{Col}(A) = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}\right\}$$

since

$$-4\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

The rowspace of A is $\text{Row}(A) = \text{Span}\left\{ \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix} \right\}$.

To determine the left nullspace of A, we solve $A^T\vec{x} = \vec{0}$. Row reducing A^T gives

$$\begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Hence, $\text{Null}(A^T) = \{\vec{0}\}$.

EXERCISE 3 Find the four fundamental subspaces of $A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ -2 & -5 & -2 & 1 \\ -1 & -3 & -1 & 2 \end{bmatrix}$

THEOREM 8 Let A be an $m \times n$ matrix. If $\vec{a} \in \text{Row}(A)$ and $\vec{x} \in \text{Null}(A)$, then $\vec{a} \cdot \vec{x} = 0$.

Proof: Let $A = \begin{bmatrix} \vec{a}_1^T \\ \vdots \\ \vec{a}_m^T \end{bmatrix}$. If $\vec{x} \in \text{Null}(A)$, then

$$\vec{0} = A\vec{x} = \begin{bmatrix} \vec{a}_1^T \\ \vdots \\ \vec{a}_m^T \end{bmatrix} \vec{x} = \begin{bmatrix} \vec{a}_1 \cdot \vec{x} \\ \vdots \\ \vec{a}_m \cdot \vec{x} \end{bmatrix}$$

So, $\vec{a}_i \cdot \vec{x} = 0$ for all $1 \leq i \leq m$. Now, if $\vec{a} \in \text{Row}(A)$, then $\vec{a} = c_1\vec{a}_1 + \cdots + c_m\vec{a}_m$. But then

$$\vec{a} \cdot \vec{x} = (c_1\vec{a}_1 + \cdots + c_m\vec{a}_m) \cdot \vec{x} = c_1(\vec{a}_1 \cdot \vec{x}) + \cdots + c_m(\vec{a}_m \cdot \vec{x}) = 0$$

as required. $\square$

THEOREM 9 Let A be an $m \times n$ matrix. If $\vec{a} \in \text{Col}(A)$ and $\vec{x} \in \text{Null}(A^T)$, then $\vec{a} \cdot \vec{x} = 0$.

Proof: The proof is left as an exercise.

3.4 Operations on Linear Mappings

Addition and Scalar Multiplication of Mappings

Since linear mappings are functions, we can perform operations on linear mappings in the same way we do any other function.

DEFINITION

Addition and Scalar Multiplication

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ and $M : \mathbb{R}^n \to \mathbb{R}^m$ be linear mappings and let $c \in \mathbb{R}$. We define $L + M : \mathbb{R}^n \to \mathbb{R}^m$ and $cL : \mathbb{R}^n \to \mathbb{R}^m$ by

$$(L + M)(\vec{x}) = L(\vec{x}) + M(\vec{x})$$
$$(cL)(\vec{x}) = cL(\vec{x})$$

REMARK

Two linear mappings L and M are equal if and only if they have the same domain, the same range, and $L(\vec{x}) = M(\vec{x})$ for all $\vec{x}$ in the domain. We write $L = M$.

EXAMPLE 1

Let $L : \mathbb{R}^2 \to \mathbb{R}^3$ and $M : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $L(x_1, x_2) = (x_1, x_1 + x_2, -x_2)$ and $M(x_1, x_2) = (x_1, x_1, x_2)$. Then, $L + M$ is the mapping defined by

$$(L + M)(\vec{x}) = L(\vec{x}) + M(\vec{x}) = (x_1, x_1 + x_2, -x_2) + (x_1, x_1, x_2) = (2x_1, 2x_1 + x_2, 0)$$

and $3L$ is the mapping defined by

$$(3L)(\vec{x}) = 3L(\vec{x}) = 3(x_1, x_1 + x_2, -x_2) = (3x_1, 3x_1 + 3x_3, -3x_2)$$

If we let $\mathbb{L}$ denote the set of all possible linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$, then we can show that $\mathbb{L}$ satisfies the same 10 properties as vectors in $\mathbb{R}^n$ and matrices in $M_{m\times n}(\mathbb{R})$.

THEOREM 1

Let $L, M, N \in \mathbb{L}$ and let c_1, c_1 be real scalars. Then

V1 $L + M \in \mathbb{L}$;

V2 $(L + M) + N = L + (M + N)$;

V3 $L + M = M + L$;

V4 There exists a linear mapping $O : \mathbb{R}^n \to \mathbb{R}^m$, such that $L+O = L$. In particular, O is the linear mapping defined by $O(\vec{x}) = \vec{0}$ for all $\vec{x} \in \mathbb{R}^n$;

V5 For each linear mapping L, there exists a linear mapping $(-L)$ with the property that $L + (-L) = O$. In particular, $(-L)(\vec{x})$ is the linear mapping defined by $(-L)(\vec{x}) = -L(\vec{x})$;

V6 $c_1 L \in \mathbb{L}$;

V7 $c_1(c_2 L) = (c_1 c_2)L$;

V8 $(c_1 + c_2)L = c_1 L + c_2 L$;

V9 $c_1(L + M) = c_1 L + c_1 M$;

V10 $1L = L$;

Proof: We will prove V1, V3, and V9 and leave the rest as exercises.

For V1, we need to show that $L + M$ is also a linear mapping. Let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $b, c \in \mathbb{R}$. Then

$$\begin{aligned}
(L+M)(b\vec{x}+c\vec{y}) &= L(b\vec{x}+c\vec{y}) + M(b\vec{x}+c\vec{y}) && \text{by definition of } L+M \\
&= bL(\vec{x}) + cL(\vec{y}) + bM(\vec{x}) + cM(\vec{y}) && \text{since } L \text{ and } M \text{ are linear} \\
&= b\big(L(\vec{x}) + M(\vec{x})\big) + c\big(L(\vec{y}) + M(\vec{y})\big) && \text{by properties of vectors in } \mathbb{R}^m \\
&= b(L+M)(\vec{x}) + c(L+M)(\vec{y}) && \text{by definition of } L+M
\end{aligned}$$

Hence, $L + M$ is linear.

For V3, we need to show $(L + M)(\vec{x}) = (M + L)(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$. We have

$$\begin{aligned}
(L+M)(\vec{x}) &= L(\vec{x}) + M(\vec{x}) && \text{by definition of } L+M \\
&= M(\vec{x}) + L(\vec{x}) && \text{using properties of vectors } \mathbb{R}^m \\
&= (M+L)(\vec{x}) && \text{by definition of } M+L
\end{aligned}$$

For V9, we need to show that $\big(c_1(L + M)\big)(\vec{x}) = \big(c_1 L + c_1 M\big)(\vec{x})$ for all $\vec{x} \in \mathbb{R}^n$. We have

$$\begin{aligned}
\big(c_1(L+M)\big)(\vec{x}) &= c_1(L+M)(\vec{x}) && \text{by definition of scalar multiplication} \\
&= c_1\big(L(\vec{x}) + M(\vec{x})\big) && \text{by definition of } L+M \\
&= c_1 L(\vec{x}) + c_1 M(\vec{x}) && \text{using properties of vectors } \mathbb{R}^m \\
&= (c_1 L)(\vec{x}) + (c_1 M)(\vec{x}) && \text{by definition of scalar multiplication} \\
&= \big(c_1 L + c_1 M\big)(\vec{x}) && \text{by definition of } L+M
\end{aligned}$$

$\square$

Properties V1 and V6 show that the sum of linear mappings is a linear mapping as is the scalar multiple of a linear mapping. Thus, it makes sense to consider the standard matrix of these new linear mappings.

THEOREM 2

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ and $M : \mathbb{R}^n \to \mathbb{R}^m$ be linear mappings and let $c \in \mathbb{R}$. Then

$$[L + M] = [L] + [M]$$
$$[cL] = c[L]$$

Proof: We have

$$[L + M](\vec{x}) = (L + M)(\vec{x}) = L(\vec{x}) + M(\vec{x}) = [L]\vec{x} + [M]\vec{x} = ([L] + [M])(\vec{x})$$

The result follows from Theorem 1. Similarly,

$$[cL](\vec{x}) = (cL)(\vec{x}) = cL(\vec{x}) = c[L]\vec{x}$$

$\square$

Composition

One of the most important operations on functions is composition.

DEFINITION
Composition

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ and $M : \mathbb{R}^m \to \mathbb{R}^p$ be linear mappings. Then M composed of L is the function $M \circ L : \mathbb{R}^n \to \mathbb{R}^p$ defined by

$$(M \circ L)(\vec{x}) = M\big(L(\vec{x})\big)$$

REMARK

Observe that it is necessary that the range of L is a subset of the domain of M for $M \circ L$ to be defined.

THEOREM 3

Let $L : \mathbb{R}^n \to \mathbb{R}^m$ and $M : \mathbb{R}^m \to \mathbb{R}^p$ be linear mappings. Then $M \circ L$ is a linear mapping and

$$[M \circ L] = [M][L]$$

Proof: The proof is left as an exercise.

Chapter 4

Vector Spaces

4.1 Vector Spaces

We have seen that $\mathbb{R}^n$, subspaces of $\mathbb{R}^n$, the set $M_{m\times n}(\mathbb{R})$ of $m \times n$ matrices, and the set $\mathbb{L}$ of all possible linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$ satisfy the same ten properties. In fact, as we will see, there are many other sets with operations of addition and scalar multiplication that satisfy these same properties. Instead of studying each of these separately, it makes sense to define and study one abstract concept which contains all of them.

DEFINITION

Vector Space

Let $\mathbb{V}$ be a set. We will call the elements of $\mathbb{V}$ **vectors** and denote them with the usual vector symbol $\vec{x}$. $\mathbb{V}$ is then called a **vector space over** $\mathbb{R}$ if there is an operation of addition, denoted $\vec{x} + \vec{y}$, and an operation of scalar multiplication, denoted $c\vec{x}$, such that for any $\vec{v}, \vec{x}, \vec{y} \in \mathbb{V}$ and $a, b \in \mathbb{R}$ we have:

- V1 $\vec{x} + \vec{y} \in \mathbb{V}$
- V2 $(\vec{x} + \vec{y}) + \vec{v} = \vec{x} + (\vec{y} + \vec{v})$
- V3 $\vec{x} + \vec{y} = \vec{y} + \vec{x}$
- V4 There is a vector denoted $\vec{0}$ in $\mathbb{V}$ such that $\vec{x} + \vec{0} = \vec{x}$. It is called the **zero vector**.
- V5 For each $\vec{x} \in \mathbb{V}$ there exists an element $-\vec{x}$ such that $\vec{x} + (-\vec{x}) = \vec{0}$. $-\vec{x}$ is called the **additive inverse** of $\vec{x}$.
- V6 $a\vec{x} \in \mathbb{V}$
- V7 $a(b\vec{x}) = (ab)\vec{x}$
- V8 $(a + b)\vec{x} = a\vec{x} + b\vec{x}$
- V9 $a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}$
- V10 $1\vec{x} = \vec{x}$.

REMARKS

1. We sometimes denote addition by $\vec{x} \oplus \vec{y}$ and scalar multiplication by $c \odot \vec{x}$ to distinguish these from "normal" operations of addition and scalar multiplication. See Example 7 below.

2. Sometimes, when working with multiple vector spaces, we denote the zero vector in a vector space $\mathbb{V}$ by $\vec{0}_{\mathbb{V}}$.

3. The ten vector space axioms define a "structure" for the set based on the operations of addition and scalar multiplication. The study of vector spaces is the study of this structure. Note that individual vector spaces may have additional operations that are not included in this structure and hence may be different than what is included in other vector spaces. For example, matrix multiplication.

4. The definition of a vector space makes sense if the scalars are allowed to be taken from any **field**, for example, $\mathbb{Q}$ or $\mathbb{C}$. In Math 235 we will consider the case where the scalars are allowed to be complex. In this course, when we say vector space, we mean a vector space over $\mathbb{R}$.

EXAMPLE 1

In Chapter 1 we saw that $\mathbb{R}^n$ and any subspace of $\mathbb{R}^n$ under standard addition and scalar multiplication of vectors satisfy all ten vector space axioms, and hence are vector spaces.

EXAMPLE 2

The set $M_{m\times n}(\mathbb{R})$ is a vector space under standard addition and scalar multiplication of matrices.

EXAMPLE 3

The set $\mathbb{L}$ of all linear mappings $L : \mathbb{R}^n \to \mathbb{R}^m$, is a vector space under standard addition and scalar multiplication of linear mappings.

EXAMPLE 4

Denote the set of all polynomials of degree at most n with real coefficients by $P_n(\mathbb{R})$. That is,

$$P_n(\mathbb{R}) = \{a_0 + a_1x + \cdots + a_nx^n \mid a_0, a_1, \ldots, a_n \in \mathbb{R}\}$$

Define standard addition and scalar multiplication of polynomials by

$$\begin{aligned}(a_0 + \cdots + a_nx^n) + (b_0 + \cdots + b_nx^n) &= (a_0 + b_0) + \cdots + (a_n + b_n)x^n \\ c(a_0 + \cdots + a_nx^n) &= (ca_0) + \cdots + (ca_n)x^n\end{aligned}$$

for all $c \in \mathbb{R}$. It is easy to verify that $P_n(\mathbb{R})$ is a vector space under these operations.

We find that the zero vector in $P_n(\mathbb{R})$ is the polynomial $z(x) = 0$ for all x, and the additive inverse of $p(x) = a_0 + a_1x + \cdots + a_nx^n$ is $(-p)(x) = -a_0 - a_1x - \cdots - a_nx^n$.

EXAMPLE 5 The set $C[a,b]$ of all continuous functions defined on the interval $[a,b]$ is a vector space under standard addition and scalar multiplication of functions.

EXAMPLE 6 Let $\mathbb{V} = \{\vec{0}\}$ and define addition by $\vec{0}+\vec{0}=\vec{0}$, and scalar multiplication by $c\vec{0}=\vec{0}$ for all $c \in \mathbb{R}$. Show that $\mathbb{V}$ is a vector space under these operations.

Solution: To show that $\mathbb{V}$ is a vector space, we need to show that it satisfies all ten axioms.

V1 The only element in $\mathbb{V}$ is $\vec{0}$ and $\vec{0}+\vec{0}=\vec{0}\in\mathbb{V}$.

V2 $(\vec{0}+\vec{0})+\vec{0}=\vec{0}+\vec{0}=\vec{0}+(\vec{0}+\vec{0})$

V3 $\vec{0}+\vec{0}=\vec{0}=\vec{0}+\vec{0}$

V4 $\vec{0}$ satisfies $\vec{0}+\vec{0}=\vec{0}$, so $\vec{0}$ is the zero vector in the set.

V5 The additive inverse of $\vec{0}$ is $\vec{0}$.

V6 $a\vec{0}=\vec{0}\in\mathbb{V}$.

V7 $a(b\vec{0})=a\vec{0}=\vec{0}=(ab)\vec{0}$.

V8 $(a+b)\vec{0}=\vec{0}=\vec{0}+\vec{0}=a\vec{0}+b\vec{0}$.

V9 $a(\vec{0}+\vec{0})=a\vec{0}=\vec{0}=\vec{0}+\vec{0}=a\vec{0}+a\vec{0}$

V10 $1\vec{0}=\vec{0}$.

for all $a,b\in\mathbb{R}$. Thus, $\mathbb{V}$ is a vector space. It is called the **trivial vector space**.

EXAMPLE 7 Let $\mathbb{S}=\{x\in\mathbb{R} \mid x>0\}$. Define addition in $\mathbb{S}$ by $x\oplus y=xy$, and define scalar multiplication by $c\odot x=x^c$ for all $x,y\in\mathbb{S}$ and all $c\in\mathbb{R}$. Prove that $\mathbb{V}$ is a vector space under these operations.

Solution: We will show that $\mathbb{S}$ satisfies all ten vector space axioms. For any $x,y,z\in\mathbb{S}$ and $a,b\in\mathbb{R}$ we have

V1 $x\oplus y=xy>0$ since $x>0$ and $y>0$. Hence, $x\oplus y\in\mathbb{S}$.

V2 $(x\oplus y)\oplus z=(xy)\oplus z=(xy)z=x(yz)=x\oplus(yz)=z\oplus(y\oplus z)$

V3 $x\oplus y=xy=yx=y\oplus x$

V4 The zero vector is 1 since $1\in\mathbb{S}$ and $x\oplus 1=x1=x$.

V5 If $x\in\mathbb{S}$, then $\frac{1}{x}\oplus x=\frac{1}{x}x=1$. Hence $\frac{1}{x}$ is the additive inverse of x. Also, $\frac{1}{x}>0$ since $x>0$ and hence $\frac{1}{x}\in\mathbb{S}$.

V6 $a \odot x = x^a > 0$ since $x > 0$. Hence $a \odot x \in \mathbb{S}$.

V7 $a \odot (b \odot x) = a \odot x^b = (x^b)^a = x^{ab} = (ab) \odot x$.

V8 $(a+b) \odot x = x^{a+b} = x^a x^b = x^a \oplus x^b = a \odot x \oplus b \odot x$.

V9 $a \odot (x \oplus y) = a \odot (xy) = (xy)^a = x^a y^a = x^a \oplus y^a = a \odot x \oplus a \odot y$.

V10 $1x = x^1 = x$.

Therefore $\mathbb{S}$ is a vector space.

EXAMPLE 8 Let $\mathbb{D} = \left\{ \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \mid a_1, a_2 \in \mathbb{R} \right\}$. Show that $\mathbb{D}$ is a vector space under standard addition and scalar multiplication of matrices.

Solution: We show that $\mathbb{D}$ satisfies all ten vector space axioms. For any $A = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$, $B = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}$, and $C = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}$ in $\mathbb{D}$ and $s, t \in \mathbb{R}$ we have

V1 $A + B = \begin{bmatrix} a_1 + b_1 & 0 \\ 0 & a_2 + b_2 \end{bmatrix} \in \mathbb{D}$.

V2 $(A + B) + C = A + (B + C)$ by Theorem 3.1.1.

V3 $A + B = B + A$ by Theorem 3.1.1.

V4 Clearly $O_{2,2} \in \mathbb{D}$ and we have $O_{2,2}$ satisfies $A + O_{2,2} = A$ for all $A \in \mathbb{D}$ by Theorem 1. Hence, $O_{2,2}$ is the zero vector in $\mathbb{D}$.

V5 By Theorem 3.1.1, the additive inverse of A is $\begin{bmatrix} -a_1 & 0 \\ 0 & -a_2 \end{bmatrix} \in \mathbb{D}$ for all $A \in \mathbb{D}$.

V6 $sA = \begin{bmatrix} sa_1 & 0 \\ 0 & sa_2 \end{bmatrix} \in \mathbb{D}$.

V7 $s(t(A)) = (st)A$ by Theorem 3.1.1.

V8 $(s + t)A = sA + tA$ by Theorem 3.1.1.

V9 $s(A + B) = sA + sB$ by Theorem 3.1.1.

V10 $1A = A$ by Theorem 3.1.1.

Hence, $\mathbb{D}$ is a vector space.

REMARK

Notice that we did not really need to check axioms V2, V3, V5, V7, V8, V9, and V10 since we were using the addition and scalar multiplication operations of a vector space. We will look more at this below.

EXAMPLE 9

Show the set $\mathbb{S} = \{(x, y) \mid x, y \in \mathbb{R}\}$ with addition defined by $(x_1, y_1) \oplus (x_2, y_2) = (x_1 + y_2, x_2 + y_1)$ and scalar multiplication defined by $c \odot (x_1, y_1) = (cx_1, cy_1)$ is not a vector space.

Solution: To show that $\mathbb{V}$ is not a vector space, we only need to find one counter example to one of the vector space axioms. Observe that $(1, 2) \in \mathbb{S}$ and that

$$\begin{aligned}(1+1) \odot (1,2) &= 2 \odot (1,2) = (2,4) \\ 1 \odot (1,2) \oplus 1 \odot (1,2) &= (1,2) \oplus (1,2) = (1+2, 2+1) = (3,3)\end{aligned}$$

Hence, axiom V8 is not satisfied. Thus, $\mathbb{S}$ is not a vector space.

EXAMPLE 10

Show the set $\mathbb{Z}^2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{Z} \right\}$ is not a vector space under standard addition and scalar multiplication of vectors.

Solution: Observe that $\begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{Z}^2$, but $\sqrt{2}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 2\sqrt{2} \end{bmatrix} \notin \mathbb{Z}^2$. Hence, $\mathbb{Z}^2$ does not satisfy axiom V6, so it is not a vector space.

We have now seen that there are numerous different vector spaces. The advantage to having all of these belonging to the one abstract concept of a vector space is that we can study all of them at the same time. That is, when we prove a result for a vector space, it is immediately true for all vector spaces. We now demonstrate this.

THEOREM 1

Let $\mathbb{V}$ be a vector space with addition defined by $\vec{x} + \vec{y}$ and scalar multiplication defined by $c\vec{x}$ for all $\vec{x}, \vec{y} \in \mathbb{V}$, and $c \in \mathbb{R}$. Then,

(1) $0\vec{x} = \vec{0}$ for all $\vec{x} \in \mathbb{V}$,
(2) $-\vec{x} = (-1)\vec{x}$ for all $\vec{x} \in \mathbb{V}$.

Proof: We prove (1) and leave (2) as an exercise. For all $\vec{x} \in \mathbb{V}$ we have

$$\begin{aligned}0\vec{x} &= 0\vec{x} + \vec{0} && \text{by V4} \\ &= 0\vec{x} + [\vec{x} + (-\vec{x})] && \text{by V5} \\ &= 0\vec{x} + [1\vec{x} + (-\vec{x})] && \text{by V10} \\ &= [0\vec{x} + 1\vec{x}] + (-\vec{x}) && \text{by V2}\end{aligned}$$

$$
\begin{aligned}
&= (0+1)\vec{x} + (-\vec{x}) && \text{by V8} \\
&= 1\vec{x} + (-\vec{x}) && \text{operation of numbers in } \mathbb{R} \\
&= \vec{x} + (-\vec{x}) && \text{by V10} \\
&= \vec{0} && \text{by V5}
\end{aligned}
$$

□

We can now use this result to find the zero vector in any vector space $\mathbb{V}$ by simply finding $0\vec{x}$ for any $\vec{x} \in \mathbb{V}$. Similarly, we can find the additive inverse of any $\vec{x} \in \mathbb{V}$ by calculating $(-1)\vec{x}$.

EXAMPLE 11

Let $\mathbb{S}$ be defined as in Example 7. Observe that we indeed have $\vec{0} = 0 \odot x = x^0 = 1$, and $(-x) = (-1) \odot x = x^{-1} = \frac{1}{x}$ for any $x \in \mathbb{S}$.

Subspaces

Observe that the set $\mathbb{D}$ in Example 8 is a vector space that is contained in the larger vector space $M_{2\times 2}(\mathbb{R})$. This is exactly the same as we saw in Chapter 1 with subspaces of $\mathbb{R}^n$.

DEFINITION
Subspace

Let $\mathbb{V}$ be a vector space. If $\mathbb{S}$ is a subset of $\mathbb{V}$ and $\mathbb{S}$ is a vector space under the same operations as $\mathbb{V}$, then $\mathbb{S}$ is called a **subspace** of $\mathbb{V}$.

As in Example 8, we see that we do not need to test all ten vector space axioms to show that a set $\mathbb{S}$ is a subspace of a vector space $\mathbb{V}$. In particular, we get the following theorem which matches what we saw in Chapter 1.

THEOREM 2

(Subspace Test)

Let $\mathbb{S}$ be a non-empty subset of $\mathbb{V}$. If $\vec{x} + \vec{y} \in \mathbb{S}$ and $c\vec{x} \in \mathbb{S}$ for all $\vec{x}, \vec{y} \in \mathbb{S}$ and $c \in \mathbb{R}$ under the operations of $\mathbb{V}$, then $\mathbb{S}$ is a subspace of $\mathbb{V}$.

Proof: The proof is left as an exercise.

REMARKS

1. It is very important to always remember to check that $\mathbb{S}$ is a subset of $\mathbb{V}$ and is non-empty before applying the subspace test.
2. As before, we will say that $\mathbb{S}$ is **closed under addition** if $\vec{x}+\vec{y} \in \mathbb{S}$ for all $\vec{x},\vec{y} \in \mathbb{S}$, and we will say that $\mathbb{S}$ is **closed under scalar multiplication** if $c\vec{x} \in \mathbb{S}$ for all $\vec{x} \in \mathbb{S}$ and $c \in \mathbb{R}$.

EXAMPLE 12

Determine which of the following are subspaces of the given vector space.

a) $\mathbb{S}_1 = \{xp(x) \mid p(x) \in P_2(\mathbb{R})\}$ of $P_4(\mathbb{R})$.

Solution: Recall that $P_2(\mathbb{R})$ is the set of all polynomials of degree at most 2. That is, polynomials of the form $p(x) = a_0 + a_1x + a_2x^2$. Similarly, $P_4(\mathbb{R})$ is the set of polynomials of degree at most 4. Then every vector $q(x) \in \mathbb{S}_1$ has the form

$$q(x) = x(a_0 + a_1x + a_2x^2) = a_0x + a_1x^2 + a_2x^3$$

which is a polynomial of degree at most 4 (actually, of at most 3). Hence, $\mathbb{S}_1$ is a subset of $P_4(\mathbb{R})$.

$\mathbb{S}_1$ is clearly non-empty since $x(1+x+x^2) \in \mathbb{S}_1$.

Let $q_1(x), q_2(x) \in \mathbb{S}_1$. Then there exist $p_1(x), p_2(x) \in P_2(\mathbb{R})$ such that $q_1(x) = xp_1(x)$ and $q_2(x) = xp_2(x)$. We have

$$q_1(x) + q_2(x) = xp_1(x) + xp_2(x) = x(p_1(x) + p_2(x))$$

and $p_1(x) + p_2(x) \in P_2(\mathbb{R})$ as $P_2(\mathbb{R})$ is closed under addition. So, $q_1(x) + q_2(x) \in \mathbb{S}_1$.

Similarly, for any $c \in \mathbb{R}$ we get

$$cq_1(x) = c(xp_1(x)) = x(cp_1(x)) \in \mathbb{S}_1$$

since $cp_1(x) \in P_2(\mathbb{R})$ as $P_2(\mathbb{R})$ is closed under scalar multiplication.

Therefore, $\mathbb{S}_1$ is a subspace of $P_4(\mathbb{R})$ by the Subspace Test.

b) $\mathbb{S}_2 = \left\{ \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix} \mid a_1a_2a_3 = 0 \right\}$ of $M_(2\times 2)(\mathbb{R})$.

Solution: Observe that $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ are both in $\mathbb{S}_2$, but

$$A + B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \notin \mathbb{S}_2$$

So, $\mathbb{S}_2$ is not closed under addition and hence is not a subspace.

c) $\mathbb{S}_3 = \{p(x) \in P_2(\mathbb{R}) \mid p(2) = 0\}$ of $P_2(\mathbb{R})$.

Solution: By definition $\mathbb{S}_3$ is a subset of $P_2(\mathbb{R})$.

In $P_2(\mathbb{R})$ the zero vector is the polynomial that satisfies $z(x) = 0$ for all x. Hence, $z(x) \in \mathbb{S}_3$ since $z(2) = 0$. Therefore, $\mathbb{S}_3$ is non-empty.

Let $p(x), q(x) \in \mathbb{S}_3$. Then $p(2) = 0$ and $q(2) = 0$, so $(p+q)(2) = p(2)+q(2) = 0+0 = 0$. Hence, $(p+q) \in \mathbb{S}_3$. Thus, $\mathbb{S}_3$ is closed under addition.

Similarly, $(cp)(2) = cp(2) = c(0) = 0$ for all $c \in \mathbb{R}$, so $(cp) \in \mathbb{S}_3$. Thus, it is also closed under scalar multiplication. Therefore $\mathbb{S}_3$ is a subspace of $P_2(\mathbb{R})$ by the Subspace Test.

d) $\mathbb{S}_4 = \left\{ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \right\}$ of $M_{2\times 2}(\mathbb{R})$.

Solution: Observe that $0 \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \notin S_3$. Therefore, $\mathbb{S}_3$ is not closed under scalar multiplication and hence is not a subspace.

REMARK

As with subspaces in Chapter 1, normally the easiest way to verify the set $\mathbb{S}$ is non-empty is to check if the zero vector of $\mathbb{V}$, $\vec{0}_{\mathbb{V}}$, is in $\mathbb{S}$. Moreover, if $\vec{0}_{\mathbb{V}} \notin \mathbb{S}$, then, as we saw in the example above, $\mathbb{S}$ is not closed under scalar multiplication and hence is not a subspace.

EXAMPLE 13

A matrix A is called **skew-symmetric** if $A^T = -A$. Determine if the set of all $n \times n$ skew-symmetric matrices with real entries is a vector space under standard addition and scalar multiplication of matrices.

Solution: To prove the set $\mathbb{S}$ of skew-symmetric matrices is a vector space, we will prove that it is a subspace of $M_{n\times n}(\mathbb{R})$.

By definition $\mathbb{S}$ is a subset of $M_{n\times n}(\mathbb{R})$. Also, the $n \times n$ zero matrix $O_{n,n}$ clearly satisfies $A^T = -A$, thus $O_{n,n} \in S$.

Let A and B be any two skew-symmetric matrices. Then

$$(A+B)^T = A^T + B^T = (-A) + (-B) = -A - B = -(A+B)$$

so $A+B$ is skew-symmetric and hence $\mathbb{S}$ is closed under addition.

Let $c \in \mathbb{R}$ then $(cA)^T = c(A^T) = c(-A) = -(cA)$ so (cA) is skew-symmetric and therefore $\mathbb{S}$ is closed under scalar multiplication.

Thus, $\mathbb{S}$ is a subspace of $M_{n\times n}(\mathbb{R})$ by the Subspace Test and therefore is a vector space.

Axioms V1 and V6 imply that every vector space is closed under linear combinations. Therefore, it makes sense to look at the concepts of spanning and linear independence as we did in Chapter 1.

Spanning

DEFINITION
Span

Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a set of vectors in a vector space $\mathbb{V}$. Then we define the **span** of $\mathcal{B}$ by

$$\operatorname{Span}\mathcal{B} = \{c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k \mid c_1, \dots, c_k \in \mathbb{R}\}$$

We say that $\operatorname{Span}\mathcal{B}$ is **spanned** by $\mathcal{B}$ and that $\mathcal{B}$ is a **spanning set** for $\operatorname{Span}\mathcal{B}$.

EXAMPLE 14

Let $\mathcal{B} = \{1 + 2x + 3x^2, 3 + 2x + x^2, 5 + 6x + 7x^2\}$. Determine which of the following vectors is in $\operatorname{Span}\mathcal{B}$.

a) $p(x) = 1 + x + 2x^2$.

Solution: To determine if $p \in \operatorname{Span}\mathcal{B}$, we need to find whether there exist coefficients c_1, c_2, and c_3 such that

$$\begin{aligned} 1 + x + 2x^2 &= c_1(1 + 2x + 3x^2) + c_2(3 + 2x + x^2) + c_3(5 + 6x + 7x^2) \\ &= (c_1 + 3c_2 + 5c_3) + (2c_1 + 2c_2 + 6c_3)x + (3c_1 + c_2 + 7c_3)x^2 \end{aligned}$$

Comparing coefficients of powers of x on both sides gives us the system of linear equations

$$\begin{aligned} c_1 + 3c_2 + 5c_3 &= 1 \\ 2c_1 + 2c_2 + 6c_3 &= 1 \\ 3c_1 + c_2 + 7c_3 &= 2 \end{aligned}$$

Row reducing the augmented matrix of the system we get

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & 1 \\ 2 & 2 & 6 & 1 \\ 3 & 1 & 7 & 2 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 3 & 5 & 1 \\ 0 & -4 & -4 & -1 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Hence, the system is inconsistent, so $p \notin \operatorname{Span}\mathcal{B}$.

b) $q(x) = 2 + x$.

Solution: To determine if $q \in \operatorname{Span}\mathcal{B}$, we need to find whether there exist coefficients c_1, c_2, and c_3 such that

$$\begin{aligned} 2 + x &= c_1(1 + 2x + 3x^2) + c_2(3 + 2x + x^2) + c_3(5 + 6x + 7x^2) \\ &= (c_1 + 3c_2 + 5c_3) + (2c_1 + 2c_2 + 6c_3)x + (3c_1 + c_2 + 7c_3)x^2 \end{aligned}$$

Comparing coefficients of powers of x on both sides gives us the system of linear equations

$$\begin{aligned} c_1 + 3c_2 + 5c_3 &= 2 \\ 2c_1 + 2c_2 + 6c_3 &= 1 \\ 3c_1 + c_2 + 7c_3 &= 0 \end{aligned}$$

Row reducing the augmented matrix of the system we get

$$\left[\begin{array}{ccc|c} 1 & 3 & 5 & 2 \\ 2 & 2 & 6 & 1 \\ 3 & 1 & 7 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 2 & -1/4 \\ 0 & 1 & 1 & 3/4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Hence, the system is consistent, so $q \in \operatorname{Span} \mathcal{B}$. In particular, if we solve the system, we get

$$\vec{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -1/4 \\ 3/4 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix}$$

for any $t \in \mathbb{R}$. Therefore, there are infinitely many different ways to write q as a linear combination of the vectors in $\mathcal{B}$.

EXAMPLE 15

Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} \right\}$. Prove that $\operatorname{Span} \mathcal{B} = \left\{ \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix} \mid a, b, c \in \mathbb{R} \right\}$.

Solution: We need to show that every vector of the form $\begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix}$ can be written as a linear combination of the vectors in $\mathcal{B}$. That is, we need to show that the equation

$$c_1 \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} + c_3 \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix}$$

is consistent for all $a_1, a_2, a_3 \in \mathbb{R}$.

Using operations on matrices on the left hand side, we get

$$\begin{bmatrix} c_1 + c_2 + c_3 & c_1 + 2c_2 + 3c_3 \\ 0 & c_1 - c_2 + 2c_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix}$$

Comparing entries gives the system of linear equations

$$\begin{aligned} c_1 + c_2 + c_3 &= a_1 \\ c_1 + 2c_2 + 3c_3 &= a_2 \\ c_1 - c_2 + 2c_3 &= a_3 \end{aligned}$$

We row reduce the coefficient matrix to get

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

By Theorem 2.2.3, the system is consistent for all $a_1, a_2, a_3 \in \mathbb{R}$.

REMARK

Notice in the last example that to prove the system was consistent for all $a_1, a_2, a_3 \in \mathbb{R}$ we only needed to row reduce the coefficient matrix. However, if we had row reduced the augmented matrix, we would have found equations for c_1, c_2, and c_3 in terms of a_1, a_2, and a_3.

THEOREM 3 Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_k\}$ be a set of vectors in a vector space $\mathbb{V}$. Then, $\operatorname{Span}\mathcal{B}$ is a subspace of $\mathbb{V}$.

Proof: The proof is left as an exercise.

Linear Independence

DEFINITION

Linearly Dependent

Linearly Independent

A set of vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ in a vector space $\mathbb{V}$ is said to be **linearly dependent** if there exist coefficients $c_1, \dots, c_k$ not all zero such that

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$$

A set of vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ is said to be **linearly independent** if the only solution to

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$$

is the trivial solution $c_1 = c_2 = \cdots = c_k = 0$.

EXAMPLE 16 Determine which of the following sets are linearly independent. If a set is linearly dependent write all linear combinations of the vectors which equals $\vec{0}$.

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

Solution: By definition, the set is linearly independent if and only if the only solution to the equation

$$0 = c_1(1+x+2x^2)+c_2(x-x^2)+c_3(-2+x^2) = (c_1-2c_3)+(c_1+c_2)x+(2c_1-c_2+c_3)x^2$$

is $c_1 = c_2 = c_3 = 0$. Comparing coefficients of powers of x gives the homogeneous system of equations

$$\begin{aligned} c_1 - 2c_3 &= 0 \\ c_1 + c_2 &= 0 \\ 2c_1 - c_2 + c_3 &= 0 \end{aligned}$$

Row reducing the coefficient matrix gives

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

By Theorem 2.2.3, the system has a unique solution. Thus, the set is linearly independent.

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

Solution: Consider

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = c_1 \begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix} + c_2 \begin{bmatrix} 1 & -2 \\ 2 & 2 \end{bmatrix} + c_3 \begin{bmatrix} 2 & 0 \\ 3 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Simplifying the right hand side gives

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 + 2c_3 + c_4 & 2c_1 - 2c_2 + c_4 \\ c_1 + 2c_2 + 3c_3 + c_4 & -2c_1 + 2c_2 + c_4 \end{bmatrix}$$

Comparing entries we get the homogeneous system

$$\begin{aligned} c_1 + c_2 + 2c_3 + c_4 &= 0 \\ 2c_1 - 2c_2 + c_4 &= 0 \\ c_1 + 2c_2 + 3c_3 + c_4 &= 0 \\ -2c_1 + 2c_2 + c_4 &= 0 \end{aligned}$$

We row reduce the coefficient matrix to get

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

Therefore, by Theorem 2.2.3, the system has infinitely many solution and hence the set is linearly dependent. In particular, finding the general solution to the system shows us that

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = (-t) \begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix} + (-t) \begin{bmatrix} 1 & -2 \\ 2 & 2 \end{bmatrix} + t \begin{bmatrix} 2 & 0 \\ 3 & 0 \end{bmatrix} + 0 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

for all $t \in \mathbb{R}$.

THEOREM 4 Any set of vectors $\{\vec{v}_1, \dots, \vec{v}_k\}$ in a vector space $\mathbb{V}$ such that $\vec{v}_i = \vec{0}$ for some i is linearly dependent.

Proof: The proof is left as an exercise.

4.2 Bases and Dimension

Bases

Theorem 4.1.3 shows that every set spanned by a set of vectors is a vector space. It is easy to see that the converse is also true. However, as we saw in Chapter 1, most of the time we wish to have a minimal spanning set for a vector space. That is, a set containing as few vectors as possible that spans the vector space. We get the following results.

THEOREM 1

If $\vec{v}_i \in \text{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$, then

$$\text{Span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \text{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$$

Proof: The proof is left as an exercise.

THEOREM 2

Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_k\}$. If $\vec{v}_i \notin \text{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$ for $1 \leq i \leq k$, then $\mathcal{B}$ is linearly independent.

Proof: We prove the contrapositive. Assume that $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly dependent. Then, there exist $c_1, \ldots, c_k$ not all zero such that

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \vec{0}$$

Pick one of the coefficients that is non-zero, say c_i. Then, we get

$$\begin{aligned} c_i\vec{v}_i &= -c_1\vec{v}_1 - \cdots - c_{i-1}\vec{v}_{i-1} - c_{i+1}\vec{v}_{i+1} - \cdots - c_k\vec{v}_k \\ \vec{v}_i &= -\frac{c_1}{c_i}\vec{v}_1 - \cdots - \frac{c_{i-1}}{c_i}\vec{v}_{i-1} - \frac{c_{i+1}}{c_i}\vec{v}_{i+1} - \cdots - \frac{c_k}{c_i}\vec{v}_k \end{aligned}$$

Hence, $\vec{v}_i \in \text{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_k\}$. □

These theorems show that a minimal spanning set for a vector space is one that is linearly independent. Thus, we make the following definition.

DEFINITION
Basis

Let $\mathbb{V}$ be a vector space. The set $\mathcal{B}$ is called a **basis** for $\mathbb{V}$ if $\mathcal{B}$ is a linearly independent spanning set for $\mathbb{V}$.

EXAMPLE 1 Prove that $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$.

Solution: For it to be a basis it must be a linearly independent spanning set. Consider

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = c_1 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3c_1 + 2c_2 + c_3 \\ 2c_1 + c_2 \\ c_1 + 5c_2 + c_3 \end{bmatrix}$$

Comparing corresponding entries we get the system of linear equations

$$\begin{aligned} 3c_1 + 2c_2 + c_3 &= b_1 \\ 2c_1 + c_2 + \qquad &= b_2 \\ c_1 + 5c_2 + c_3 &= b_3 \end{aligned}$$

Row reducing the corresponding coefficient matrix we get

$$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 5 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since there is a leading 1 in each row, the system is consistent for every $\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \in \mathbb{R}^3$ by Theorem 2.2.3. Hence $\mathcal{B}$ spans $\mathbb{R}^3$.

Similarly, since there is a leading 1 in each column, the homogeneous system ($b_1 = b_2 = b_3 = 0$) has a unique solution by Theorem 2.2.3. Thus $\mathcal{B}$ is also linearly independent.

Therefore, $\mathcal{B}$ is a basis for $\mathbb{R}^3$.

EXAMPLE 2 The set $\{1, x, x^2, \ldots, x^n\}$ is a basis for $P_n(\mathbb{R})$ since it is clearly linearly independent and a spanning set for $P_n(\mathbb{R})$. It is called the standard basis for $P_n(\mathbb{R})$.

EXAMPLE 3 The set $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is the standard basis for $M_{2\times 2}(\mathbb{R})$.

EXAMPLE 4 Prove that $\mathcal{B} = \{1 + 2x + x^2, 2 + 9x, 3 + 3x + 4x^2\}$ is a basis for $P_2(\mathbb{R})$.

Solution: First, we need to show that for any polynomial $p(x) = c + bx + ax^2$ we can find c_1, c_2, c_3 such that

$$\begin{aligned} c + bx + ax^2 &= c_1(1 + 2x + x^2) + c_2(2 + 9x) + c_3(3 + 3x + 4x^2) \\ &= (c_1 + 2c_2 + 3c_3) + (2c_1 + 9c_2 + 3c_3)x + (c_1 + 4c_3)x^2. \end{aligned}$$

Row reducing the coefficient matrix of the corresponding system gives

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 3 \\ 1 & 0 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

By Theorem 2.2.3, Span $\mathcal{B} = P_2(\mathbb{R})$ since the system is consistent for every right hand side $\begin{bmatrix} c \\ b \\ a \end{bmatrix}$. Moreover, if we let $a = b = c = 0$, we see that we get the unique solution $c_1 = c_2 = c_3 = 0$, and hence $\mathcal{B}$ is also linearly independent. Therefore, it is a basis for $P_2(\mathbb{R})$.

EXAMPLE 5

Determine whether $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} \right\}$ is a basis for the vector space

$$\mathbb{U} = \left\{ \begin{bmatrix} a & b \\ 0 & c \end{bmatrix} \mid a, b, c \in \mathbb{R} \right\}$$

Solution: Consider

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

This gives us the homogeneous system of linear equations

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ c_1 + 2c_2 + 3c_3 &= 0 \\ c_1 - c_2 + 2c_3 &= 0 \end{aligned}$$

Row reducing the coefficient matrix of the homogeneous system we get

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & -1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence, the only solution is $c_1 = c_2 = c_3 = 0$, so the set is linearly independent.

Now, consider

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

This gives a system of linear equations with the same coefficient matrix as above. So, using the same row-operations we did above we see that the matrix has a leading one in each row of its REF, hence the system is consistent by Theorem 2.2.3 for all matrices $\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$, so the set spans $\mathbb{U}$.

Therefore, $\mathcal{B}$ is a basis for $\mathbb{U}$.

EXAMPLE 6 Find a basis for the vector space $\mathbb{S}$ of 2×2 skew-symmetric matrices.

Solution: We first want to find a spanning set for $\mathbb{S}$. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is skew-symmetric, then $A^T = -A$, so

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}$$

For this to be true, we must have $a = -a$, $c = -b$, $b = -c$, and $d = -d$. Thus, $a = 0 = d$ and $b = -c$. Hence, every vector in $\mathbb{S}$ can be written in the form

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix} = b \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

Thus, $\mathbb{S} = \operatorname{Span}\left\{ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$, so $\mathcal{B} = \left\{ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$ is a spanning set for $\mathbb{S}$. Moreover, $\mathcal{B}$ contains one non-zero vector, so it is clearly linearly independent. Therefore, $\mathcal{B}$ is a basis for $\mathbb{S}$.

In Linear Algebra we often wish to find a basis for a vector space. The example above shows the general procedure for finding a basis for a vector space which can be spanned by a finite number of vectors.

ALGORITHM

Let $\mathbb{V}$ be a vector space. To find a basis for $\mathbb{V}$:

1. Find the general form of a vector $\vec{x} \in \mathbb{V}$.
2. Write the general form of $\vec{x}$ as a linear combination of vectors $\vec{v}_1, \ldots, \vec{v}_k$ in $\mathbb{V}$.
3. Check if $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent. If it is, then stop as $\mathcal{B}$ is a basis.
4. If $\mathcal{B}$ is linearly dependent, pick some vector $\vec{v}_i \in \mathcal{B}$ which can be written as a linear combination of the other vectors in $\mathcal{B}$ and remove it from the set using Theorem 1. Repeat this until $\mathcal{B}$ is linearly independent.

EXAMPLE 7 Find a basis for the vector space $\mathbb{V} = \left\{ \begin{bmatrix} a-b \\ a-c \\ c-b \end{bmatrix} \mid a, b, c \in \mathbb{R} \right\}$.

Solution: Any vector $\vec{v} \in \mathbb{V}$ has the form

$$\vec{v} = \begin{bmatrix} a-b \\ a-c \\ c-b \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} + c \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Thus, $\mathbb{V} = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \right\}$. Consider

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

Row reducing the coefficient matrix of the corresponding system of linear equations gives

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, the set is linearly dependent. However, interpreting the coefficient matrix above as an augmented matrix, we get that

$$\left[\begin{array}{cc|c} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right]$$

This tells us that

$$\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = - \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}$$

So, we can apply Theorem 1, to get that the set $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} \right\}$ also spans $\mathbb{V}$.

Now consider

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 - c_2 \\ c_1 \\ -c_2 \end{bmatrix}$$

Using exactly the same row operations as above gives

$$\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Hence, the system has a unique solution, so $\mathcal{B}$ is linearly independent. Thus, $\mathcal{B}$ is a basis for $\mathbb{V}$.

In general, it would be very inefficient to remove one vector at a time to reduce a spanning set to a basis. Moreover, notice that when we did remove a vector from $\mathcal{B}$ in the example above, that all we did was remove the corresponding column from the matrix representation of the corresponding system of equations. So, we often can use our knowledge and understanding of solving systems of equations to be able to remove all linearly dependent vectors at once. This is demonstrated in the example below.

EXAMPLE 8

Consider the vectors p_1, p_2, p_3, p_4, and p_5 given by

$$\begin{aligned} p_1(x) &= 4 + 7x - 12x^3 + 9x^4, & p_2(x) &= -2 + 12x - 8x^2 + 4x^4, \\ p_3(x) &= -16 + 3x - 16x^2 + 36x^3 - 19x^4, & p_4(x) &= -6 + 19x^2 - 6x^3, \\ p_5(x) &= 82 - 49x - 96x^2 - 12x^3 + 17x^4, \end{aligned}$$

in $P_4(\mathbb{R})$, and let $\mathbb{V} = \operatorname{Span}\{p_1, p_2, p_3, p_4, p_5\}$. Determine a basis for $\mathbb{V}$ given that the following two matrices are row equivalent:

$$\begin{bmatrix} 4 & -2 & -16 & -6 & 82 \\ 7 & 12 & 3 & 0 & -49 \\ 0 & -8 & -16 & 19 & -96 \\ -12 & 0 & 36 & -6 & -12 \\ 9 & 4 & -19 & 0 & 17 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & -3 & 0 & 5 \\ 0 & 1 & 2 & 0 & -7 \\ 0 & 0 & 0 & 1 & -8 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Solution: Observe that the first matrix is the coefficient matrix of the homogeneous system corresponding to the equation

$$0 = c_1p_1 + c_2p_2 + c_3p_3 + c_4p_4 + c_5p_5$$

The RREF shows us that the set $\{p_1, p_2, p_3, p_4, p_5\}$ is linearly dependent. However, observe that if we ignore the last two columns, and interpret the first three columns as an augmented matrix, then this shows that

$$p_3 = -3p_1 + 2p_2$$

Similarly, interpreting all the columns as an augmented matrix and taking $c_3 = 0$, we find that

$$p_5 = 5p_1 - 7p_2 - 8p_4$$

Hence, by Theorem 1, we get that $\mathbb{V} = \operatorname{Span}\{p_1, p_2, p_4\}$. Moreover, if we ignore the third and fifth column of the matrices, we see that the equation

$$0 = c_1p_1 + c_2p_2 + c_4p_4$$

has a unique solution, hence $\{p_1, p_2, p_4\}$ is linearly independent. Thus, $\{p_1, p_2, p_4\}$ is a basis for $\mathbb{V}$.

Dimension

In Chapter 1, we geometrically understood that $\mathbb{R}^n$ is n-dimensional and, for example, a plane in $\mathbb{R}^n$ is two-dimensional. We now generalize the concept of dimension to vector spaces. We begin with the following important theorems.

THEOREM 3 Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for $\mathbb{V}$, and let $\{\vec{w}_1, \ldots, \vec{w}_k\}$ be a linearly independent set in $\mathbb{V}$. Then $k \leq n$.

Proof: Since $\vec{w}_1, \ldots, \vec{w}_k \in \mathbb{V}$ and $\mathcal{B}$ is a basis for $\mathbb{V}$, we can write every vector $\vec{w}_1, \ldots, \vec{w}_k$ as a linear combination of the vectors in $\mathcal{B}$. Say,

$$\begin{aligned}
\vec{w}_1 &= a_{11}\vec{v}_1 + a_{12}\vec{v}_2 + \cdots + a_{1n}\vec{v}_n \\
\vec{w}_2 &= a_{21}\vec{v}_1 + a_{22}\vec{v}_2 + \cdots + a_{2n}\vec{v}_n \\
\vdots \quad & \quad \vdots \qquad\qquad\qquad\qquad \vdots \\
\vec{w}_k &= a_{k1}\vec{v}_1 + a_{k2}\vec{v}_2 + \cdots + a_{kn}\vec{v}_n
\end{aligned}$$

Consider,

$$\begin{aligned}
\vec{0} &= c_1\vec{w}_1 + \cdots + c_k\vec{w}_k \\
&= c_1(a_{11}\vec{v}_1 + a_{12}\vec{v}_2 + \cdots + a_{1n}\vec{v}_n) + \cdots + c_k(a_{k1}\vec{v}_1 + a_{k2}\vec{v}_2 + \cdots + a_{kn}\vec{v}_n) \\
&= (c_1a_{11} + \cdots + c_ka_{k1})\vec{v}_1 + \cdots + (c_1a_{1n} + \cdots + c_ka_{kn})\vec{v}_n
\end{aligned}$$

Since $\mathcal{B}$ is a basis, it is linearly independent, and hence the only solution is

$$\begin{aligned}
c_1a_{11} + \cdots + c_ka_{k1} &= 0 \\
\vdots \qquad\qquad \vdots \quad & \quad \vdots \\
c_1a_{1n} + \cdots + c_ka_{kn} &= 0
\end{aligned}$$

If $k > n$, then the system has infinitely many solutions by Theorem 2.2.3. But, this would imply that $\vec{0} = c_1\vec{w}_1 + \cdots + c_k\vec{w}_k$ has infinitely many solutions which contradicts our assumption that $\{\vec{w}_1, \ldots, \vec{w}_k\}$ is linearly independent. Hence, $k \leq n$. $\square$

REMARK

Recall that we showed that we can think of a basis as a minimal spanning set. This theorem shows that we can also think of a basis as a maximal linearly independent set. That is, a basis $\mathcal{B}$ is a linearly independent set such that if we add any other vector from the vector space to $\mathcal{B}$, then the resulting set would be linearly dependent.

THEOREM 4 Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ and $\{\vec{w}_1, \ldots, \vec{w}_k\}$ be bases for a vector space $\mathbb{V}$. Then, $k = n$.

Proof: The proof is left as an exercise.

Thus, if a vector space $\mathbb{V}$ has a basis containing a finite number of elements, then every basis of $\mathbb{V}$ has the same number of elements. This implies that every basis for $\mathbb{R}^n$ contains exactly n vectors, and every basis for a plane in $\mathbb{R}^n$ contains exactly two vectors which matches our geometric understanding of dimension. Thus, we make the following definition.

DEFINITION

Dimension

Let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for a vector space $\mathbb{V}$. Then, we say the **dimension** of $\mathbb{V}$ is n and we write

$$\dim V = n$$

REMARK

If $\mathbb{V}$ does not have a basis with a finite number of vectors in it, then $\mathbb{V}$ is said to be infinite-dimensional. We define a basis for the trivial vector space to be the empty set, so that the dimension of the trivial vector space is 0.

EXAMPLE 9

The dimension of $\mathbb{R}^n$ is n since the standard basis contains n vectors.

The dimension of $M_{m\times n}(\mathbb{R})$ is mn since the standard basis contains mn vectors.

The dimension of $P_n(\mathbb{R})$ is $n+1$ since the standard basis contains $n+1$ vectors.

EXAMPLE 10

Since $\left\{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right\}$ is a basis for the vector space $\mathbb{S}$ of 2×2 skew-symmetric matrices, we have that the dimension of $\mathbb{S}$ is 1.

EXAMPLE 11

In Example 7, we found that a basis for $\mathbb{V} = \left\{\begin{bmatrix} a-b \\ a-c \\ c-b \end{bmatrix} \mid a, b, c \in \mathbb{R}\right\}$ is $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}\right\}$. Thus, $\dim \mathbb{V} = 2$. So, geometrically $\mathbb{V}$ is a plane in $\mathbb{R}^3$.

The concept of dimension can be very useful in many cases. This is demonstrated in the following theorem.

THEOREM 5

Let $\mathbb{V}$ be an n-dimensional vector space. Then

(1) A set of more than n vectors in $\mathbb{V}$ must be linearly dependent.
(2) A set of fewer than n vectors in $\mathbb{V}$ cannot span $\mathbb{V}$.
(3) A set of n vectors in $\mathbb{V}$ is linearly independent if and only if it spans $\mathbb{V}$.

Proof: The proof is left as an exercise.

REMARK

Observe that (3) shows that if we have a set $\mathcal{B}$ of n linearly independent vectors in an n-dimensional vector space $\mathbb{V}$, then $\mathcal{B}$ is a basis for $\mathbb{V}$. Similarly, if $\mathcal{C}$ contains n vectors and $\operatorname{Span}\mathcal{C} = \mathbb{V}$, then $\mathcal{C}$ is a basis for $\mathbb{V}$.

EXAMPLE 12

Find a basis for the hyperplane $x_1 + x_2 + 2x_3 + x_4 = 0$ in $\mathbb{R}^4$ and then extend the basis to obtain a basis for $\mathbb{R}^4$.

Solution: By definition of a hyperplane, we know that a hyperplane in $\mathbb{R}^4$ has dimension 3. Hence, by Theorem 5, any three linearly independent vectors in the hyperplane will form a basis for the hyperplane. Observe that the vectors $\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 0 \\ 2 \\ -1 \\ 0 \end{bmatrix}$, and $\vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -2 \end{bmatrix}$ satisfy the equation of the hyperplane and hence are in the hyperplane. Now, consider

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 = \begin{bmatrix} c_1 \\ -c_1 + 2c_2 \\ -c_2 + c_3 \\ -2c_3 \end{bmatrix}$$

The only solution is $c_1 = c_2 = c_3 = 0$ so the vectors are linearly independent. Thus, $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a basis for the hyperplane.

To extend the basis $\mathcal{B}$ to a basis for $\mathbb{R}^4$, we just need to add a vector $\vec{v}_4$ so that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is linearly independent. Since $\vec{v}_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ does not satisfy the equation of the hyperplane, it is not in $\operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ and so is not a linear combination of $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Thus, $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is linearly independent and therefore a basis for $\mathbb{R}^4$.

THEOREM 6

Let $\mathbb{V}$ be an n-dimensional vector space, and let $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be a linearly independent set in $\mathbb{V}$ with $k < n$. Then, there exist vectors $\vec{w}_{k+1}, \ldots, \vec{w}_n$ in $\mathbb{V}$ such that $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{w}_{k+1}, \ldots, \vec{w}_n\}$ is a basis for $\mathbb{V}$.

Proof: The proof is left as an exercise.

We will use this fact that we can always extend a linearly independent set to a basis for a finite dimensional vector space in a few proofs in Math 235.

4.3 Coordinates

Observe that when we write a vector $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^n$, that we really mean

$$\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$$

where $\{\vec{e}_1, \dots, \vec{e}_n\}$ is the standard basis for $\mathbb{R}^n$. For example, when you originally learned to plot the point $(1, 2)$ in the xy-plane, you were taught that this means you move 1 in the x-direction ($\vec{e}_1$) and 2 in the y-direction ($\vec{e}_2$). Of course, we can extend this to general vector spaces.

Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for a vector space $\mathbb{V}$. Then every vector $\vec{v}$ in $\mathbb{V}$ can be written as a linear combination of the vectors in $\mathcal{B}$. Say, $\vec{v} = b_1\vec{v}_1 + \cdots + b_n\vec{v}_n$. Thus, we can think of $\vec{v}$ as being b_1 in the $\vec{v}_1$ direction, b_2 in the $\vec{v}_2$ direction, etc. Therefore, with respect to the basis $\mathcal{B}$, it is the coefficients of the linear combination of the basis vectors which gives $\vec{v}$ that defines the vector $\vec{v}$. This is demonstrated in the next example.

EXAMPLE 1

Let $\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} 0 \\ 5 \\ -4 \end{bmatrix}$, $p(x) = 1 - 2x + 3x^2$, $q(x) = 5x - 4x^2$. Also, let $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ be a basis for $\mathbb{V} = \operatorname{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and let $\vec{v} = \vec{v}_1 - 2\vec{v}_2 + 3\vec{v}_3$, and $\vec{w} = 0\vec{v}_1 + 5\vec{v}_2 - 4\vec{v}_3$. Then we have

$$\vec{x} + \vec{y} = \begin{bmatrix} 1+0 \\ -2+5 \\ 3-4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$$
$$p(x) + q(x) = (1+0) + (-2+5)x + (3-4)x^2 = 1 + 3x - x^2$$
$$\vec{v} + \vec{w} = (1+0)\vec{v}_1 + (-2+5)\vec{v}_2 + (3-4)\vec{v}_3 = 1\vec{v}_1 + 3\vec{v}_2 - \vec{v}_3$$

From this example we notice that once we have written out the vectors with respect to a basis, performing operations on vectors in any finite dimensional vector space is exactly the same as operations on vectors in $\mathbb{R}^n$. However, before we proceed any further, we show that this concept is well defined by showing that if we have a basis for a vector space, then every vector can be written as a *unique* linear combination of the basis vectors.

THEOREM 1

Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for $\mathbb{V}$. Then, every vector $\vec{v} \in \mathbb{V}$ can be represented as a *unique* linear combination of $\vec{v}_1, \dots, \vec{v}_n$.

Proof: Since $\mathcal{B}$ is a basis, it is a spanning set. Thus, for every $\vec{v} \in \mathbb{V}$ there exist $c_1, \dots, c_n$ such that

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

Assume that there also exists $d_1, \ldots, d_n$ such that $d_1\vec{v}_1 + \cdots + d_n\vec{v}_n = \vec{v}$. Then

$$c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = d_1\vec{v}_1 + \cdots + d_n\vec{v}_n \Rightarrow (c_1 - d_1)\vec{v}_1 + \cdots + (c_n - d_n)\vec{v}_n = \vec{0}$$

But, this implies that $c_i = d_i$ for $1 \leq i \leq n$ since $\mathcal{B}$ is linearly independent. Thus, there is only one linear combination of the vectors in $\mathcal{B}$ that equals $\vec{v}$. $\square$

This allows us to make the following definition.

DEFINITION

Coordinate Vector

Let $\mathbb{V}$ be a vector space with basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$. For any $\vec{v} \in \mathbb{V}$ we define the **coordinate vector** of $\vec{v}$ with respect to $\mathcal{B}$ by

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

where $\vec{v} = b_1\vec{v}_1 + \cdots + b_n\vec{v}_n$.

REMARK

We will often say the **$\mathcal{B}$-coordinates** of a vector rather than saying the coordinate vector of $\vec{v}$ with respect to $\mathcal{B}$.

EXAMPLE 2

Given that the set $\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \right\}$ is a basis for the subspace of $M_{2\times 2}(\mathbb{R})$ which it spans, find the $\mathcal{B}$-coordinates of $\vec{u} = \begin{bmatrix} 1 & -3 \\ 2 & 3 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} -1 & 0 \\ 3 & 7 \end{bmatrix}$.

Solution: We need to find c_1, c_2, c_3 and d_1, d_2, d_3 such that

$$\begin{aligned} c_1 \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} + c_2 \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} &= \begin{bmatrix} 1 & -3 \\ 2 & 3 \end{bmatrix} \\ d_1 \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} + d_2 \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + d_3 \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} &= \begin{bmatrix} -1 & 0 \\ 3 & 7 \end{bmatrix} \end{aligned}$$

Observe that each corresponding system will have the same coefficient matrix, but different augmented part. Thus, we can solve both systems simultaneously by row reducing a doubly augmented matrix. We get

$$\left[\begin{array}{ccc|cc} 1 & 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & -3 & 0 \\ 0 & 1 & 1 & 2 & 3 \\ -1 & 1 & 2 & 3 & 7 \end{array}\right] \sim \left[\begin{array}{ccc|cc} 1 & 0 & 0 & -2 & -2 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

For the first system we have $c_1 = -2$, $c_2 = 3$, and $c_3 = -1$, and for the second system we have $d_1 = -2$, $d_2 = 1$, and $d_3 = 2$. Thus,

$$[\vec{u}]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} \quad \text{and} \quad [\vec{v}]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$$

REMARK

It is important to notice that the coordinates of a vector is dependent on which basis is being used for the vector space. Moreover, it is also dependent on the order of vectors in the basis. Thus, to prevent confusion, when we speak of a basis, we mean an ordered basis.

As intended, coordinate vectors allow us to change from working with vectors in some vector space with respect to some basis to working with vectors in $\mathbb{R}^n$. In particular, observe that taking coordinates of a vector $\vec{v}$ with respect to a basis $\mathcal{B}$ of a vector space $\mathbb{V}$ is really a function which takes a vector in $\mathbb{V}$ and outputs a vector in $\mathbb{R}^n$. Not surprisingly, this function is linear.

THEOREM 2 Let $\mathbb{V}$ be a vector space with basis $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$. Then, for any $\vec{v}, \vec{w} \in \mathbb{V}$ and $s, t \in \mathbb{R}$ we have

$$[s\vec{v} + t\vec{w}]_{\mathcal{B}} = s[\vec{v}]_{\mathcal{B}} + t[\vec{w}]_{\mathcal{B}}$$

Proof: Let $\vec{v} = b_1\vec{v}_1 + \cdots + b_n\vec{v}_n$ and $\vec{w} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n$. Then we have

$$s\vec{v} + t\vec{w} = (sb_1 + tc_1)\vec{v}_1 + \cdots + (sb_n + tc_n)\vec{v}_n$$

Therefore,

$$[s\vec{v} + t\vec{w}]_{\mathcal{B}} = \begin{bmatrix} sb_1 + tc_1 \\ \vdots \\ sb_n + tc_n \end{bmatrix} = s\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} + t\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = s[\vec{v}]_{\mathcal{B}} + t[\vec{w}]_{\mathcal{B}}$$

$\square$

This function also satisfies two other important properties: it is one-to-one and onto. As we will see in Math 235, this will allow us to show that, as vector spaces, any n-dimensional vector space is essentially the same as $\mathbb{R}^n$.

Change of Coordinates

We have seen how to find different bases for a vector space $\mathbb{V}$ and how to find the coordinates of a vector $\vec{v} \in \mathbb{V}$ with respect to any basis $\mathcal{B}$ of $\mathbb{V}$. In some cases, it is useful to have a quick way of determining the coordinates of $\vec{v}$ with respect to some basis $\mathcal{C}$ for $\mathbb{V}$ given the coordinates of $\vec{v}$ with respect to the basis $\mathcal{B}$.

For example, in some applications it is useful to write polynomials in terms of powers of $x - c$. That is, given any polynomial $p(x) = a_0 + a_1x + \cdots + a_nx^n$, you want to write it as

$$p(x) = b_0 + b_1(x - c) + b_2(x - c)^2 + \cdots + b_n(x - c)^n$$

Such a situation may arise if the values of x you are working with are very close to c. If you are working with many polynomials, then it would be very helpful to have a fast way of converting each polynomial. We can rephrase this problem in terms of linear algebra.

Let $\mathcal{S} = \{1, x, \dots, x^n\}$ be the standard basis for $P_n(\mathbb{R})$. Then, given $[p(x)]_\mathcal{S} = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}$ we want to determine $[p(x)]_\mathcal{B}$ where $\mathcal{B}$ is the basis $\mathcal{B} = \{1, x-c, (x-c)^2, \dots, (x-c)^n\}$ for $P_n(\mathbb{R})$.

Observe that since taking coordinates is a linear operation by Theorem 2, we get

$$\begin{aligned}
[p(x)]_\mathcal{B} &= [a_0 1 + a_1 x + \cdots + a_n x^n]_\mathcal{B} \\
&= a_0[1]_\mathcal{B} + a_1[x]_\mathcal{B} + \cdots + a_n[x^n]_\mathcal{B} \\
&= \begin{bmatrix} [1]_\mathcal{B} & [x]_\mathcal{B} & \cdots & [x^n]_\mathcal{B} \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}
\end{aligned}$$

So, the multiplication of this matrix and the coordinate vector of $p(x)$ with respect to the standard basis gives us the coordinates of $p(x)$ with respect to the new basis $\mathcal{B}$. We call this matrix the **change of coordinates matrix** from $\mathcal{S}$ coordinates to $\mathcal{B}$ coordinates and denote it by ${}_\mathcal{B}P_\mathcal{S}$.

EXAMPLE 3 Consider the basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ -3 \end{bmatrix} \right\}$ for $\mathbb{R}^3$. Find the change of coordinates matrix from $\mathcal{B}$ to the standard basis $\mathcal{S}$.

Solution: The columns of the desired change of coordinates matrix are the $\mathcal{S}$-coordinates of the vectors in $\mathcal{B}$. We have

$$\left[\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right]_\mathcal{S} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \qquad \left[\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right]_\mathcal{S} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \qquad \left[\begin{bmatrix} -2 \\ 0 \\ -3 \end{bmatrix} \right]_\mathcal{S} = \begin{bmatrix} -2 \\ 0 \\ -3 \end{bmatrix}$$

So, the change of coordinates matrix ${}_{\mathcal{S}}P_{\mathcal{B}}$ is

$$ {}_{\mathcal{S}}P_{\mathcal{B}} = \begin{bmatrix} \left[\begin{bmatrix}1\\2\\3\end{bmatrix}\right]_{\mathcal{S}} & \left[\begin{bmatrix}-1\\0\\1\end{bmatrix}\right]_{\mathcal{S}} & \left[\begin{bmatrix}-2\\0\\-3\end{bmatrix}\right]_{\mathcal{S}} \end{bmatrix} = \begin{bmatrix}1 & -1 & -2\\ 2 & 0 & 0\\ 3 & 1 & -3\end{bmatrix} $$

Of course, we can repeat the argument above in the case where $\mathcal{B}$ and $\mathcal{C}$ are any two bases of a vector space $\mathbb{V}$. In particular, if $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix}b_1\\ \vdots\\ b_n\end{bmatrix}$, then

$$ \begin{aligned} [\vec{x}]_{\mathcal{C}} &= [b_1\vec{v}_1 + \cdots + b_n\vec{v}_n]_{\mathcal{C}} \\ &= b_1[\vec{v}_1]_{\mathcal{C}} + \cdots + b_n[\vec{v}_n]_{\mathcal{C}} \\ &= \begin{bmatrix}[\vec{v}_1]_{\mathcal{C}} & \cdots & [\vec{v}_n]_{\mathcal{C}}\end{bmatrix}\begin{bmatrix}b_1\\ \vdots\\ b_n\end{bmatrix} \end{aligned} $$

Hence, we make the following definition.

DEFINITION

Change of Coordinates Matrix

Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $\mathcal{C}$ both be bases for a vector space $\mathbb{V}$. Then, the **change of coordinates matrix** from $\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates is defined by

$$ {}_{\mathcal{C}}P_{\mathcal{B}} = \begin{bmatrix}[\vec{v}_1]_{\mathcal{C}} & \cdots & [\vec{v}_n]_{\mathcal{C}}\end{bmatrix} $$

and for any $\vec{x} \in \mathbb{V}$ we have

$$ [\vec{x}]_{\mathcal{C}} = {}_{\mathcal{C}}P_{\mathcal{B}}[\vec{x}]_{\mathcal{B}} $$

EXAMPLE 4

Let $\mathcal{B} = \left\{\begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\}$ and $\mathcal{C} = \left\{\begin{bmatrix}-1\\1\end{bmatrix}, \begin{bmatrix}5\\-4\end{bmatrix}\right\}$ be bases for $\mathbb{R}^2$. Find ${}_{\mathcal{C}}P_{\mathcal{B}}$ and ${}_{\mathcal{B}}P_{\mathcal{C}}$.

Solution: To find ${}_{\mathcal{C}}P_{\mathcal{B}}$ we need to find the $\mathcal{C}$-coordinates of the vectors in $\mathcal{B}$. Thus, we need to solve the systems

$$ c_{11}\begin{bmatrix}-1\\1\end{bmatrix} + c_{12}\begin{bmatrix}5\\-4\end{bmatrix} = \begin{bmatrix}1\\3\end{bmatrix} \qquad c_{21}\begin{bmatrix}-1\\1\end{bmatrix} + c_{22}\begin{bmatrix}5\\-4\end{bmatrix} = \begin{bmatrix}2\\1\end{bmatrix} $$

Observe that since system has the same coefficient matrix, the same row operations will solve each system. Therefore, to make this shorter, we make the multiple augmented system

$$ \left[\begin{array}{cc|cc} -1 & 5 & 1 & 2\\ 1 & -4 & 3 & 1 \end{array}\right] $$

Row reducing this system gives

$$\left[\begin{array}{cc|cc} -1 & 5 & 1 & 2 \\ 1 & -4 & 3 & 1 \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & 19 & 13 \\ 0 & 1 & 4 & 3 \end{array}\right]$$

Hence, ${}_{\mathcal{C}}P_{\mathcal{B}} = \begin{bmatrix} 19 & 13 \\ 4 & 3 \end{bmatrix}$. Similarly, for ${}_{\mathcal{B}}P_{\mathcal{C}}$ we need to find the $\mathcal{B}$-coordinates of the vectors in $\mathcal{C}$. We again form a multiple augmented system and row reduce to get

$$\left[\begin{array}{cc|cc} 1 & 2 & -1 & 5 \\ 3 & 1 & 1 & -4 \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & 3/5 & -13/5 \\ 0 & 1 & -4/5 & 19/5 \end{array}\right]$$

So ${}_{\mathcal{B}}P_{\mathcal{C}} = \begin{bmatrix} 3/5 & -13/5 \\ -4/5 & 19/5 \end{bmatrix}$.

EXAMPLE 5

Let $\mathcal{B} = \{1 + x, 1 + 2x + x^2, x - x^2\}$ be a basis for P_2. Determine the change of coordinates matrix from $\mathcal{B}$ to the standard basis $\mathcal{S} = \{1, x, x^2\}$ for P_2.

Solution: We need to find the coordinates of the vectors in $\mathcal{B}$ with respect to the basis $\mathcal{S}$. We have

$$[1+x]_{\mathcal{S}} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad [1+2x+x^2]_{\mathcal{S}} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad [x-x^2]_{\mathcal{S}} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Hence, the change of coordinates matrix is

$${}_{\mathcal{S}}P_{\mathcal{B}} = \begin{bmatrix} [1+x]_{\mathcal{S}} & [1+2x+x^2]_{\mathcal{S}} & [x-x^2]_{\mathcal{S}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

EXAMPLE 6

Consider the basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \right\}$. Find the change of coordinates matrix from the standard basis $\mathcal{S}$ to $\mathcal{B}$ and hence find $\left[\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right]_{\mathcal{B}}$.

Solution: The columns of the desired change of coordinates matrix are the $\mathcal{B}$-coordinates of the standard basis vectors. To find these we need to solve the augmented systems

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 3 & 1 & 4 & 0 \\ -1 & 1 & 1 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 3 & 1 & 4 & 1 \\ -1 & 1 & 1 & 0 \end{array}\right], \quad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 3 & 1 & 4 & 0 \\ -1 & 1 & 1 & 1 \end{array}\right]$$

Again, we row reduce the multiple augmented system to get

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 3 & 1 & 4 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3/5 & -1/5 & -1 \\ 0 & 1 & 0 & 7/5 & -4/5 & -1 \\ 0 & 0 & 1 & -4/5 & 3/5 & 1 \end{array}\right]$$

Thus, the change of coordinates matrix from $\mathcal{S}$-coordinates to $\mathcal{B}$-coordinates is

$$_{\mathcal{B}}P_{\mathcal{S}} = \begin{bmatrix} 3/5 & -1/5 & -1 \\ 7/5 & -4/5 & -1 \\ -4/5 & 3/5 & 1 \end{bmatrix}$$

Therefore, we have

$$\begin{aligned} \left[\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right]_{\mathcal{B}} &= \begin{bmatrix} 3/5 & -1/5 & -1 \\ 7/5 & -4/5 & -1 \\ -4/5 & 3/5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{3}{5}x_1 - \frac{1}{5}x_2 - x_3 \\ \frac{7}{5}x_1 - \frac{4}{5}x_2 - x_3 \\ -\frac{4}{5}x_1 + \frac{3}{5}x_2 + x_3 \end{bmatrix} \end{aligned}$$

We can check to make sure that the answer in Example 6 is correct by verifying that these are the $\mathcal{B}$-coordinates of $[\vec{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Observe that if $\left[\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right]_{\mathcal{B}} = \begin{bmatrix} \frac{3}{5}x_1 - \frac{1}{5}x_2 - x_3 \\ \frac{7}{5}x_1 - \frac{4}{5}x_2 - x_3 \\ -\frac{4}{5}x_1 + \frac{3}{5}x_2 + x_3 \end{bmatrix}$, then by definition of coordinates we have

$$\begin{aligned} (\frac{3}{5}x_1 - \frac{1}{5}x_2 - x_3) \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} + (\frac{7}{5}x_1 - \frac{4}{5}x_2 - x_3) \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + (-\frac{4}{5}x_1 + \frac{3}{5}x_2 + x_3) \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \\ = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 4 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{3}{5}x_1 - \frac{1}{5}x_2 - x_3 \\ \frac{7}{5}x_1 - \frac{4}{5}x_2 - x_3 \\ -\frac{4}{5}x_1 + \frac{3}{5}x_2 + x_3 \end{bmatrix} \\ = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 4 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3/5 & -1/5 & -1 \\ 7/5 & -4/5 & -1 \\ -4/5 & 3/5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned}$$

Hence, we not only see that our calculations were correct, but we observe that since

${}_{\mathcal{S}}P_{\mathcal{B}} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 4 \\ -1 & 1 & 1 \end{bmatrix}$ we have shown that

$${}_{\mathcal{S}}P_{\mathcal{B}}\,{}_{\mathcal{B}}P_{\mathcal{S}} = I$$

Of course, we can prove this in general.

THEOREM 3 Let $\mathcal{B}$ and $\mathcal{C}$ be bases for an n-dimensional vector space $\mathbb{V}$. Then the change of coordinate matrices ${}_{\mathcal{C}}P_{\mathcal{B}}$ and ${}_{\mathcal{B}}P_{\mathcal{C}}$ satisfy

$${}_{\mathcal{C}}P_{\mathcal{B}}\,{}_{\mathcal{B}}P_{\mathcal{C}} = I = {}_{\mathcal{B}}P_{\mathcal{C}}\,{}_{\mathcal{C}}P_{\mathcal{B}}$$

Proof: By definition of the change of coordinates matrix we have $[\vec{x}]_{\mathcal{C}} = {}_{\mathcal{C}}P_{\mathcal{B}}[\vec{x}]_{\mathcal{B}}$ and $[\vec{x}]_{\mathcal{B}} = {}_{\mathcal{B}}P_{\mathcal{C}}[\vec{x}]_{\mathcal{C}}$. Combining these gives

$$[\vec{x}]_{\mathcal{C}} = {}_{\mathcal{C}}P_{\mathcal{B}}[\vec{x}]_{\mathcal{B}} = {}_{\mathcal{C}}P_{\mathcal{B}}\,{}_{\mathcal{B}}P_{\mathcal{C}}[\vec{x}]_{\mathcal{C}}$$

Hence $I[\vec{x}]_{\mathcal{C}} = {}_{\mathcal{C}}P_{\mathcal{B}}\,{}_{\mathcal{B}}P_{\mathcal{C}}[\vec{x}]_{\mathcal{C}}$ for all $[\vec{x}]_{\mathcal{C}} \in \mathbb{R}^n$, so ${}_{\mathcal{C}}P_{\mathcal{B}}\,{}_{\mathcal{B}}P_{\mathcal{C}} = I$ by Theorem 4. Similarly,

$$[\vec{x}]_{\mathcal{B}} = {}_{\mathcal{B}}P_{\mathcal{C}}[\vec{x}]_{\mathcal{C}} = {}_{\mathcal{B}}P_{\mathcal{C}}\,{}_{\mathcal{C}}P_{\mathcal{B}}[\vec{x}]_{\mathcal{B}}$$

for all $[\vec{x}]_{\mathcal{B}} \in \mathbb{R}^n$, hence ${}_{\mathcal{B}}P_{\mathcal{C}}\,{}_{\mathcal{C}}P_{\mathcal{B}} = I$. $\square$

Chapter 5

Inverses and Determinants

5.1 Matrix Inverses

In the last chapter we saw that if $\mathcal{B}$ and $\mathcal{C}$ are bases for a finite dimensional vector space $\mathbb{V}$, then the change of coordinates matrix ${}_{\mathcal{B}}P_{\mathcal{C}}$ and the change of coordinates matrix ${}_{\mathcal{C}}P_{\mathcal{B}}$ satisfy

$$ {}_{\mathcal{B}}P_{\mathcal{C}}\,{}_{\mathcal{C}}P_{\mathcal{B}} = I $$

Since the product of these matrices gives the identity matrix (the multiplicative identity), these matrices should be inverses of each other. We now develop and explore the idea of inverse matrices.

DEFINITION

Left and Right Inverse

Let A be an $m \times n$ matrix. If B is an $n \times m$ matrix such that $AB = I_m$, then B is called the **right inverse** of A. If C is an $n \times m$ matrix such that $CA = I_n$, then C is called the **left inverse** of A.

EXAMPLE 1

Let $A = \begin{bmatrix} 1 & 2 & 2 \\ -1 & 3 & 1 \end{bmatrix}$. A right inverse of A is $B = \begin{bmatrix} 3/5 & -2/5 \\ 1/5 & 1/5 \\ 0 & 0 \end{bmatrix}$ since

$$\begin{bmatrix} 1 & 2 & 2 \\ -1 & 3 & 1 \end{bmatrix}\begin{bmatrix} 3/5 & -2/5 \\ 1/5 & 1/5 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Note that this also shows that A is the left inverse of B.

THEOREM 1

If A is an $m \times n$ matrix with $m > n$, then A cannot have a right inverse.

Proof: The proof is left as an exercise.

COROLLARY 2

If A is an $m \times n$ matrix with $m < n$, then A cannot have a left inverse.

Proof: If A has a left inverse C, then C is an $n \times m$ matrix with $n > m$ with a right inverse which contradicts the previous theorem. □

Thus we see that only an $n \times n$ matrix can have a left and a right inverse. We now show that if a matrix has a left and right inverse, then the left and right inverses must equal.

THEOREM 3

Let A, B and C be $n \times n$ matrices such that $AB = I = CA$, then $B = C$.

Proof: We have

$$B = IB = (CA)B = C(AB) = CI = C$$

□

DEFINITION

Matrix Inverse

Invertible

Let A be an $n \times n$ matrix. If B is a matrix such that $AB = I = BA$, then B is called the **inverse** of A. We write $B = A^{-1}$ and we say that A is **invertible**.

REMARK

Observe from the definition that if $B = A^{-1}$, then $A = B^{-1}$.

The following theorem is extremely important. It not only shows that the right (left) inverse of an $n \times n$ matrix is necessarily the inverse, but also that the rank of an invertible matrix is n.

THEOREM 4

Let A be an $n \times n$ matrix. If there exists an $n \times n$ matrix B such that $AB = I$, then $\operatorname{rank} A = n = \operatorname{rank} B$ and $BA = I$. Hence, A is invertible.

Proof: The proof is left as an exercise.

THEOREM 5

Let A and B be invertible matrices and let $c \in \mathbb{R}$, $c \neq 0$. Then

(1) $(cA)^{-1} = \frac{1}{c}A^{-1}$,
(2) $(A^T)^{-1} = (A^{-1})^T$,
(3) $(AB)^{-1} = B^{-1}A^{-1}$.

Proof: We just prove (1) and leave (2) and (3) as exercises.

By Theorem 4, to show that $B = A^{-1}$ we just need to show that $AB = I$. Thus, to show that $(cA)^{-1} = B = \frac{1}{c}A^{-1}$ we just need to show that $(cA)\left(\frac{1}{c}A^{-1}\right) = I$. We have

$$(cA)\left(\frac{1}{c}A^{-1}\right) = \frac{c}{c}AA^{-1} = 1I = I$$

as required. $\square$

THEOREM 6 Let A be an $n \times n$ matrix with rank n. Then A is invertible.

Proof: Observe that if rank $A = n$, then the system of equations $A\vec{b}_i = \vec{e}_i$, $1 \leq i \leq n$ are all consistent by Theorem 2.2.3. Hence, if we let $B = \begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_n \end{bmatrix}$, then

$$AB = A\begin{bmatrix} \vec{b}_1 & \cdots & \vec{b}_n \end{bmatrix} = \begin{bmatrix} A\vec{b}_1 & \cdots & A\vec{b}_n \end{bmatrix} = \begin{bmatrix} \vec{e}_1 & \cdots & \vec{e}_n \end{bmatrix} = I$$

Thus, A is invertible by Theorem 4. $\square$

Observe that the proof of this theorem teaches us one way of finding the inverse of a matrix A. The columns of the inverse are the vectors $\vec{b}_i$ such that $A\vec{b}_i = \vec{e}_i$ for $1 \leq i \leq n$. Since each of these systems have the same coefficient matrix A, we can solve them by making one multiple augmented matrix and row reducing. In particular, assuming that A is invertible we have

$$[A \mid I] \sim \left[I \mid A^{-1}\right]$$

REMARK

Compare this with the method for finding the change of coordinates matrix from standard coordinates in $\mathbb{R}^n$ to any other basis $\mathcal{B}$ of $\mathbb{R}^n$.

EXAMPLE 2 Find the inverse of $A = \begin{bmatrix} 1 & -2 & 3 \\ -1 & 4 & 1 \\ 1 & -1 & 4 \end{bmatrix}$.

Solution: We row reduce the multiple augmented system $[A \mid I]$.

$$\left[\begin{array}{ccc|ccc} 1 & -2 & 3 & 1 & 0 & 0 \\ -1 & 4 & 1 & 0 & 1 & 0 \\ 1 & -1 & 4 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -17/2 & -5/2 & 7 \\ 0 & 1 & 0 & -5/2 & -1/2 & 2 \\ 0 & 0 & 1 & 3/2 & 1/2 & -1 \end{array}\right]$$

Thus, $A^{-1} = \begin{bmatrix} -17/2 & -5/2 & 7 \\ -5/2 & -1/2 & 2 \\ 3/2 & 1/2 & -1 \end{bmatrix}$.

EXAMPLE 3

Let $A \in M_{2\times 2}(\mathbb{R})$. Find a condition on the entries of A that guarantee that A is invertible and then, assuming that A is invertible, find A^{-1}.

Solution: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. For A to be invertible, we require that $\operatorname{rank} A = 2$. Thus, we cannot have $a = 0 = c$. So, we assume that $a \neq 0$. Row reducing $[A \mid I]$ we get

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \begin{array}{c} \\ R_2 - \frac{c}{a}R_1 \end{array} \sim \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & d - \frac{bc}{a} & -\frac{c}{a} & 1 \end{array}\right] \begin{array}{c} \\ \\ aR_2 \end{array} \sim \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & ad - bc & -c & a \end{array}\right]$$

Since we need $\operatorname{rank} A = 2$, we now require that $ad - bc \neq 0$. Assuming this we continue row reducing to get

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & ad - bc & -c & a \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

Thus, we get that A is invertible if and only if $ad - bc \neq 0$ and if A is invertible, then

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

REMARK

This quantity $ad - bc$ which determines whether a 2×2 matrix is invertible or not is called the **determinant** of the matrix. We will look more at this later in the chapter.

Invertible Matrix Theorem

The following theorem ties together many of the concepts we have looked at throughout the course. Additionally, we will find it useful throughout the rest of the course.

THEOREM 7

(Invertible Matrix Theorem)

For an $n \times n$ matrix A, the following are equivalent:

(1) A is invertible,
(2) The RREF of A is I,
(3) $\operatorname{rank} A = n$,
(4) The system of equations $A\vec{x} = \vec{b}$ is consistent with a unique solution for all $\vec{b} \in \mathbb{R}^n$,

(5) The nullspace of A is $\{\vec{0}\}$,
(6) The columns of A form a basis for $\mathbb{R}^n$,
(7) The rows of A form a basis for $\mathbb{R}^n$,
(8) A^T is invertible.

This theorem says that if any one of the eight items are true for an $n \times n$ matrix, then all of the other items are also true. We leave the proof of this theorem as an important exercise. This is an excellent way of testing that you understand much of the material we have done in the course up to this point.

The Invertible Matrix Theorem tells us that if A is invertible, then the system of equations $A\vec{x} = \vec{b}$ is consistent with a unique solution. What is the solution? Observe that since A is invertible, we have

$$\begin{aligned} A^{-1}A\vec{x} &= A^{-1}\vec{b} \\ I\vec{x} &= A^{-1}\vec{b} \\ \vec{x} &= A^{-1}\vec{b} \end{aligned}$$

Hence, if A is invertible, we can solve the system $A\vec{x} = \vec{b}$ by simply matrix multiplying $A^{-1}\vec{b}$.

EXAMPLE 4 Solve the system of linear equations

$$\begin{aligned} 3x_1 - 2x_2 &= 4 \\ x_1 + 5x_2 &= 7 \end{aligned}$$

Solution: We can represent the system as $A\vec{x} = \vec{b}$ where $A = \begin{bmatrix} 3 & -2 \\ 1 & 5 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$. Then, from our work in Example 3, we see that A is invertible with $A^{-1} = \frac{1}{17}\begin{bmatrix} 5 & 2 \\ -1 & 3 \end{bmatrix}$. Hence, the solution of the system is

$$\vec{x} = A^{-1}\vec{b} = \frac{1}{17}\begin{bmatrix} 5 & 2 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

EXAMPLE 5 Let $A = \begin{bmatrix} 1 & -2 & 3 \\ -1 & 4 & 1 \\ 1 & -1 & 4 \end{bmatrix}$. Solve the following systems of linear equations.

a) $A\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Solution: In Example 2 we found that $A^{-1} = \begin{bmatrix} -17/2 & -5/2 & 7 \\ -5/2 & -1/2 & 2 \\ 3/2 & 1/2 & -1 \end{bmatrix}$. Hence, the solution of the system is

$$\vec{x} = A^{-1}\vec{b} = \begin{bmatrix} -17/2 & -5/2 & 7 \\ -5/2 & -1/2 & 2 \\ 3/2 & 1/2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/2 \\ -1/2 \\ 1/2 \end{bmatrix}$$

b) $A\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

Solution: The solution is

$$\vec{x} = A^{-1}\vec{b} = \begin{bmatrix} -17/2 & -5/2 & 7 \\ -5/2 & -1/2 & 2 \\ 3/2 & 1/2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -13/2 \\ -3/2 \\ 3/2 \end{bmatrix}$$

Observe from the example above that once we have found A^{-1} we can solve any system of linear equations with coefficient matrix A very quickly. Moreover, at first glance, it might seem like we are solving the system of linear equations without row reducing. However, this is an illusion. The row operations are in fact "stored" inside of the inverse of the matrix (we used row operations to find the inverse). We now look at how to discover those stored row operations, and the connection between row operations and matrix multiplication.

5.2 Elementary Matrices

Consider the system of linear equations $A\vec{x} = \vec{b}$ where A is invertible. We compare our two methods of solving this system.

First, we can solve this system using the methods of Chapter 2 by row reducing the augmented matrix:

$$\left[A \mid \vec{b}\right] \sim [I \mid \vec{x}]$$

Alternately, we can solve it by using the fact that A is invertible. We find A^{-1} using the method in the previous section:

$$[A \mid I] \sim \left[I \mid A^{-1}\right]$$

We then solve the system by computing $\vec{x} = A^{-1}\vec{b}$.

Observe that in both methods we can use exactly the same row operations to row reduce A to I. In the first method, we are applying those row operations directly to

$\vec{b}$ to determine $\vec{x}$. Thus, we see that in the second method the row operations used are being "stored" inside of A^{-1} and the matrix vector product $A^{-1}\vec{b}$ is "performing" those row operations on $\vec{b}$ so that we get the same answer as in the first method.

This not only shows us that A^{-1} is made of elementary row operations, but also that there is a close connection between matrix multiplication and performing elementary row operations.

DEFINITION

Elementary Matrix

An $n \times n$ matrix E is called an **elementary matrix** if it can be obtained from the identity matrix by performing exactly one elementary row operation.

EXAMPLE 1

Determine which of the following matrices are elementary. For each elementary matrix, indicate the associated elementary row operation.

a) $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Solution: This matrix is elementary as it can be obtained from the identity matrix by performing the row operation $R_1 + 2R_3$.

b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Solution: This matrix is elementary as it can be obtained from the identity matrix by performing the row operation $R_1 \leftrightarrow R_2$.

c) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

Solution: This matrix is not elementary as it would require three elementary row operations to get this matrix from the identity matrix.

d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Solution: This matrix is elementary as it can be obtained from the identity matrix by performing the row operation $(-1)R_2$.

It is clear that the RREF of every elementary matrix is I; we just need to perform the inverse (opposite) row operation to turn the elementary matrix E back to I. Thus, we not only have that every elementary matrix is invertible, but the inverse is the elementary matrix associated with the opposite row operation.

EXAMPLE 2

Find the inverse of each of the following elementary matrices. Check your answer by multiplying the matrices together.

a) $E_1 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

Solution: The inverse matrix E_1^{-1} is the elementary matrix associated with the row operation required to bring E_1 back to I. That is, $R_1 - 2R_2$. Therefore,

$$E_1^{-1} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

Checking, we get

$$E_1^{-1}E_1 = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

b) $E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

Solution: The inverse matrix E_2^{-1} is the elementary matrix associated with the row operation required to bring E_2 back to I. That is, $R_1 \leftrightarrow R_3$. Therefore,

$$E_2^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Checking, we get

$$E_2^{-1}E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

c) $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix}$

Solution: The inverse matrix E_3^{-1} is the elementary matrix associated with the row operation required to bring E_3 back to I. That is, $(-1/3)R_3$. Therefore,

$$E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/3 \end{bmatrix}$$

Checking, we get

$$E_3^{-1}E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

If we look at our calculations in the example above we see something interesting. Consider an elementary matrix E. We saw that the product EI is the matrix obtained from I by performing the row operations associated with E on I, and that $E^{-1}E$ is the matrix obtained from E by performing the row operation associated with E^{-1} on E. We demonstrate this further with another example.

EXAMPLE 3 Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, and $A = \begin{bmatrix} 1 & 1 & 3 \\ 0 & -1 & 2 \\ 0 & 5 & 6 \end{bmatrix}$. Calculate E_1A and E_2E_1A. Describe the products in terms of matrices obtain from A by elementary row operations.

Solution: By matrix multiplication, we get

$$E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 \\ 0 & -1 & 2 \\ 0 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 1 & 8 \\ 0 & 5 & 6 \end{bmatrix}$$

Observe that E_1A is the matrix obtain from A by performing the row operation $R_2 + 2R_1$. That is, by performing the row operation associated with E_1.

$$E_2E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 \\ 2 & 1 & 8 \\ 0 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 1 & 8 \\ 0 & 15 & 18 \end{bmatrix}$$

Thus, E_2E_1A is the matrix obtained from A by first performing the row operation $R_2 + 2R_1$, and then performing the row operation associated with E_2, namely $3R_3$.

The following theorems prove that we have this useful fact for all elementary matrices.

THEOREM 1 Let A be an $m \times n$ matrix and let E be the $m \times m$ elementary matrix corresponding to the row operation $R_i + cR_j$, for $i \neq j$. Then EA is the matrix obtained from A by performing the row operation $R_i + cR_j$ on A.

Proof: Observe that the i-th row of E satisfies $e_{ii} = 1$, $e_{ij} = c$, and $e_{ik} = 0$ for all $k \neq i, j$. So, the entries in the i-th row of EA are

$$\begin{aligned} (EA)_{ik} = \sum_{\ell=1}^{m} E_{i\ell}A_{\ell k} \qquad &= e_{ij}a_{jk} + e_{ii}a_{ik} \\ = ca_{jk} + a_{ik} & \end{aligned}$$

for $1 \leq k \leq n$. Therefore, the i-th row of A is equal to $R_i + cR_j$. The remaining rows of E are rows from the identity matrix, and hence the remaining rows of EA are just the corresponding rows in A. $\square$

THEOREM 2 Let A be an $m \times n$ matrix and let E be the $m \times m$ elementary matrix corresponding to the row operation cR_i. Then EA is the matrix obtained from A by performing the row operation cR_i on A.

Proof: The proof is left as an exercise.

THEOREM 3 Let A be an $m \times n$ matrix and let E be the $m \times m$ elementary matrix corresponding to the row operation $R_i \leftrightarrow R_j$, for $i \neq j$. Then EA is the matrix obtained from A by performing the row operation $R_i \leftrightarrow R_j$ on A.

Proof: The proof is left as an exercise.

Since elementary row operations do not change the rank of a matrix, we immediately get the following result.

COROLLARY 4 Let A be an $m \times n$ matrix and let E be an $m \times m$ elementary matrix. Then,

$$\operatorname{rank}(EA) = \operatorname{rank} A$$

We began this section by asking about how the inverse of a matrix stores elementary row operations. We can now use our work with elementary matrices to decompose a matrix into its elementary components.

THEOREM 5 Let A be an $m \times n$ matrix. Then there exists a sequence $E_1, \dots, E_k$ of $m \times m$ elementary matrices such that

$$E_k E_{k-1} \cdots E_2 E_1 A = R$$

where R is the reduced row echelon form of A.

Proof: Using the methods of Chapter 2, we know that we can row reduce a matrix A to it RREF with a sequence of elementary row operations. Let E_1 denote the first row operations performed, E_2 the second row operation, etc. The result now follows from Theorems 1, 2, and 3, since $E_1 A$ is the matrix obtained from A by performing the first row operation on A, $E_2 E_1 A$ is the matrix obtained from A by performing the second row operation on $E_1 A$, etc. $\square$

EXAMPLE 4 Let $A = \begin{bmatrix} 1 & 3 & 0 & 1 \\ -1 & -3 & 3 & 2 \\ 2 & 7 & 0 & 2 \end{bmatrix}$. Write the reduced row echelon form R of A as a product of elementary matrices and A.

Solution: We first row reduce A to R keeping track of the row operations used.

$$\begin{bmatrix} 1 & 3 & 0 & 1 \\ -1 & -3 & 3 & 2 \\ 2 & 7 & 0 & 2 \end{bmatrix} \begin{matrix} \\ R_2 + R_1 \\ R_3 - 2R_1 \end{matrix} \sim \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 0 & 3 & 3 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ R_2 \leftrightarrow R_3 \end{matrix} \sim$$

$$\begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 3 \end{bmatrix} \begin{matrix} \\ \\ \frac{1}{3}R_3 \end{matrix} \sim \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{matrix} R_1 - 3R_2 \\ \\ \\ \end{matrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Next, we write the elementary matrix for each row operation we used. We have

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \qquad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$E_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \qquad E_5 = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, we have

$$E_5E_4E_3E_2E_1A = R$$

COROLLARY 6 If A is an $n \times n$ invertible matrix, then A and A^{-1} can be written as a product of elementary matrices.

Proof: Since A is invertible we have that the RREF A is I by the Invertible Matrix Theorem. Hence, by Theorem 5, we get that there exist elementary matrices $E_1, \ldots, E_k$ such that

$$E_k \cdots E_1 A = I$$

Thus, by Theorem 5.1.4 we have that

$$A^{-1} = E_k \cdots E_1$$

So, A^{-1} can be written as a product of elementary matrices. Moreover, by Theorem 5.1.5 (3), we get that

$$A = (E_k \cdots E_1)^{-1} = E_1^{-1}E_2^{-1} \cdots E_k^{-1}$$

as required. □

EXAMPLE 5 Let $A = \begin{bmatrix} 1 & 3 \\ -3 & -11 \end{bmatrix}$. Write A and A^{-1} as products of elementary matrices.

Solution: Row reducing A to RREF gives

$$\begin{bmatrix} 1 & 3 \\ -3 & -11 \end{bmatrix} \begin{matrix} \\ R_2 + 3R_1 \end{matrix} \sim \begin{bmatrix} 1 & 3 \\ 0 & -2 \end{bmatrix} \begin{matrix} \\ (-1/2)R_2 \end{matrix} \sim$$
$$\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{matrix} R_1 - 3R_2 \\ \\ \end{matrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Hence,

$$E_1 = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \quad E_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1/2 \end{bmatrix} \quad E_3 = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$$

and $E_3E_2E_1A = I$ so

$$A^{-1} = E_3E_2E_1.$$

Moreover, we have

$$A = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$$

We end this section with one additional result which will be used in a couple of proofs later in the course.

THEOREM 7 Let E be an $m \times m$ elementary matrix. Then E^T is an elementary matrix.

Proof: If E is obtained from I by multiplying a row by a non-zero constant c, then $E^T = E$.

If E is obtained from I by swapping row i and row j, then $e_{ij} = 1 = e_{ji}$, so $E^T = E$.

Finally, if E is obtained from I by adding c times row i to row j, then E^T is obtained from I by adding c times row j to row i.

Hence, in all cases, E^T is elementary. $\square$

5.3 Determinants

We have seen that a system of linear equations with coefficient matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has a unique solution if and only if $ad - bc \neq 0$. Therefore, by the Invertible Matrix Theorem, A is invertible if and only if $ad - bc \neq 0$. Since this quantity determines whether the matrix is invertible or not, we make the following definition.

DEFINITION

2×2 **Determinant**

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The **determinant** of A is

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

EXAMPLE 1

Calculate the following determinants.

a) $\det \begin{bmatrix} 3 & 5 \\ -1 & 2 \end{bmatrix}$

Solution: $\det \begin{bmatrix} 3 & 5 \\ -1 & 2 \end{bmatrix} = 3(2) - 5(-1) = 11$

b) $\begin{vmatrix} 3 & 6 \\ 2 & 4 \end{vmatrix}$

Solution: $\begin{vmatrix} 3 & 6 \\ 2 & 4 \end{vmatrix} = 3(4) - 2(6) = 0$

We, of course, want to extend the definition of the determinant to $n \times n$ matrices. So, we next consider the 3×3 case. Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. Then it can be shown (with some effort) that A is invertible if and only if

$$a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} \neq 0.$$

Thus, we define this to be the determinant of the 3×3 matrix A. However, we would like to write the formula in a nicer form. Moreover, we would like to find some sort of pattern with this and the 2×2 determinant so that we can figure out how to define the determinant for general $n \times n$ matrices. We observe that

$$\begin{aligned} &a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} \\ &\qquad = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22}) \\ &\qquad = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \end{aligned} \tag{5.1}$$

Therefore, the 3×3 determinant can be written as a linear combination of 2×2 determinants of submatrices of A where the coefficients are (plus or minus) the coefficients from the first column of A.

To extend all of this to $n \times n$ matrices we will recursively use the same pattern. We first make some definitions.

DEFINITION

Cofactor

Let A be an $n \times n$ matrix with $n \geq 2$. Let $A(i,j)$ be the $n-1 \times n-1$ matrix obtained from A by deleting the i-th row and the j-th column. The **cofactor** of a_{ij} is

$$C_{ij} = (-1)^{i+j} \det A(i,j)$$

EXAMPLE 2

Find all nine of the cofactors of $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 3 \\ -2 & -3 & 4 \end{bmatrix}$.

Solution: We have

$$\begin{aligned} C_{11} &= (-1)^{1+1} \begin{vmatrix} -1 & 3 \\ -3 & 4 \end{vmatrix} & C_{12} &= (-1)^{1+2} \begin{vmatrix} 0 & 3 \\ -2 & 4 \end{vmatrix} & C_{13} &= (-1)^{1+3} \begin{vmatrix} 0 & -1 \\ -2 & -3 \end{vmatrix} \\ &= 5 & &= -6 & &= -2 \\ C_{21} &= (-1)^{2+1} \begin{vmatrix} 0 & 2 \\ -3 & 4 \end{vmatrix} & C_{22} &= (-1)^{2+2} \begin{vmatrix} 1 & 2 \\ -2 & 4 \end{vmatrix} & C_{23} &= (-1)^{2+3} \begin{vmatrix} 1 & 0 \\ -2 & -3 \end{vmatrix} \\ &= -6 & &= 8 & &= 3 \\ C_{31} &= (-1)^{3+1} \begin{vmatrix} 0 & 2 \\ -1 & 3 \end{vmatrix} & C_{32} &= (-1)^{3+2} \begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} & C_{33} &= (-1)^{3+3} \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} \\ &= 2 & &= -3 & &= -1 \end{aligned}$$

Using this and equation (5.1) we get

$$\det A = a_{11}C_{11} + a_{21}C_{21} + a_{31}C_{31} = \sum_{i=1}^{3} a_{i1}C_{i1}.$$

Then, to define the determinant of an $n \times n$ matrix, we repeat this recursive pattern.

DEFINITION

$n \times n$ Determinant

Let A be an $n \times n$ matrix with $n \geq 2$. Then, the **determinant** of A is defined to be

$$\det A = \sum_{i=1}^{n} a_{i1}C_{i1}$$

where the determinant of a 1×1 matrix is defined to be $\det[c] = c$.

REMARK

As we did with 2×2 matrices, we often represent the determinant of a matrix by $\left| \quad \right|$. For example, we write

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

EXAMPLE 3 Find the determinant of $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ -2 & 2 & 3 \end{bmatrix}$.

Solution: By definition, we have

$$\begin{aligned} \det A &= a_{11}C_{11} + a_{21}C_{21} + a_{31}C_{31} \\ &= 1(-1)^{1+1}\begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix} + 0(-1)^{2+1}\begin{vmatrix} 0 & 2 \\ 2 & 3 \end{vmatrix} + (-2)(-1)^{3+1}\begin{vmatrix} 0 & 2 \\ 1 & 1 \end{vmatrix} \\ &= 1(1)\big(1(3) - 1(2)\big) + 0(1)\big(0(3) - 2(2)\big) + (-2)(1)\big(0(1) - 2(1)\big) \\ &= 1 + 0 + 4 = 5 \end{aligned}$$

In the example above we did not actually need to evaluate the cofactor C_{12} since the coefficient $a_{12} = 0$. Normally, such steps are omitted.

EXAMPLE 4 Find the determinant of $B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -2 & 0 & 2 & 3 \end{bmatrix}$.

Solution: By definition, we have

$$\begin{aligned} \det B &= b_{11}C_{11} + b_{21}C_{21} + b_{31}C_{31} + b_{41}C_{41} \\ &= (-1)^{2+1}\begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & 3 \end{vmatrix} + (-2)(-1)^{4+1}\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \\ &= (-1)\begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & 3 \end{vmatrix} + 2\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \end{aligned}$$

We now need to evaluate these 3×3 determinants. By definition, we get

$$\begin{aligned} \det B =& (-1)\left((-1)^{1+1}\begin{vmatrix} 0 & 1 \\ 2 & 3 \end{vmatrix} + (-1)^{2+1}\begin{vmatrix} 0 & 0 \\ 2 & 3 \end{vmatrix}\right) + \\ & 2\left((-1)^{1+1}\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + (-1)^{3+1}\begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}\right) \\ =& (-1)\Big(\big(0(3) - 1(2)\big) + (-1)\big(0(3) - 0(2)\big)\Big) + \\ & 2\Big(\big(1(1) - 0(0)\big) + \big(0(0) - 0(1)\big)\Big) \\ =& 2 + 2 = 4 \end{aligned}$$

This example clearly demonstrates the recursive nature of the definition of the determinant. In particular, notice that we were quite lucky to have lots of zeros in the matrix as otherwise we would have had many more cofactors to evaluate. In general, if we use this formula to find the determinant of an $n \times n$ matrix, we need to calculate the value of n cofactors, each of which is the determinant of an $(n-1)\times(n-1)$ matrix. To calculate each cofactor we need to calculate the value of $n-1$ determinants of $(n-2)\times(n-2)$ matrices, and so on. For even a small value of n, say $n = 1000$, you can image that this is going to take a lot of work. In some modern applications, it is necessary to calculate determinants of much, much larger matrices. Therefore, we want to search for faster ways to evaluate a determinant.

The Cofactor Expansion

Observe that our way of factoring the 3×3 determinant in equation (5.1) was not the only way it could be factored. We also could get

$$\begin{aligned}
\det A &= a_{12}C_{12} + a_{22}C_{22} + a_{32}C_{32} \\
\det A &= a_{13}C_{13} + a_{23}C_{23} + a_{33}C_{33} \\
\det A &= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \\
\det A &= a_{21}C_{21} + a_{22}C_{22} + a_{23}C_{23} \\
\det A &= a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33}
\end{aligned}$$

Hence, we did not have to use the coefficients and cofactors from the first column of the matrix. We can in fact use the coefficients and cofactors from any row or column of the matrix. For $n \times n$ matrices we have the following theorem.

THEOREM 1 Let A be an $n \times n$ matrix. Then

$$\det A = \sum_{k=1}^{n} a_{ik}C_{ik}$$

called the cofactor expansion across the i-th row, OR

$$\det A = \sum_{k=1}^{n} a_{kj}C_{kj}$$

called the cofactor expansion across the j-th column.

REMARK

The proof of this theorem is fairly lengthy and so is omitted. However, the proof is actually quite interesting, so you may wish to search for it, or, if you are interested in a challenge, try proving it yourself.

This theorem allows us to expand the determinant along the row or column with the most zeros.

EXAMPLE 5 Find the determinant of each of the following matrices.

a) $A = \begin{bmatrix} 1 & 2 & 0 \\ -3 & 4 & 6 \\ -2 & 5 & 0 \end{bmatrix}$.

Solution: Using the cofactor expansion along the third columns gives

$$\begin{aligned}
\det A &= a_{13}C_{13} + a_{23}C_{23} + a_{33}C_{33} \\
&= 6(-1)^{2+3}\begin{vmatrix} 1 & 2 \\ -2 & 5 \end{vmatrix} \\
&= -6\big(1(5) - 2(-2)\big) \\
&= -54
\end{aligned}$$

b) $B = \begin{bmatrix} 1 & 4 & 2 & 5 \\ 0 & 1 & 0 & 0 \\ -1 & 3 & 0 & 1 \\ 0 & -2 & 3 & 0 \end{bmatrix}$

Solution: Using the cofactor expansion along the second row gives

$$\begin{aligned}
\det B &= b_{21}C_{21} + b_{22}C_{22} + b_{23}C_{23} + b_{24}C_{24} \\
&= 1(-1)^{2+2}\begin{vmatrix} 1 & 2 & 5 \\ -1 & 0 & 1 \\ 0 & 3 & 0 \end{vmatrix}
\end{aligned}$$

We now expand the 3×3 determinant along the third row to get

$$\begin{aligned}
\det B &= 1(3)(-1)^{3+2}\begin{vmatrix} 1 & 5 \\ -1 & 1 \end{vmatrix} \\
&= -3\big(1(1) - 5(-1)\big) \\
&= -18
\end{aligned}$$

c) $C = \begin{bmatrix} 1 & 51 & -73 & 29 \\ 0 & 2 & -16 & 15 \\ 0 & 0 & 3 & -99 \\ 0 & 0 & 0 & 4 \end{bmatrix}$.

Solution: Using the cofactor expansion along the first column in each step gives

$$\begin{aligned}
\det C &= 1(-1)^{1+1}\begin{vmatrix} 2 & -16 & 15 \\ 0 & 3 & -99 \\ 0 & 0 & 4 \end{vmatrix} \\
&= 1(2)(-1)^{1+1}\begin{vmatrix} 3 & -99 \\ 0 & 4 \end{vmatrix} \\
&= 1(2)\big(3(4) - (-99)(0)\big) \\
&= 1(2)(3)(4) = 24
\end{aligned}$$

d) $D = \begin{bmatrix} 3 & -3 & 4 & 0 & 5 \\ -6 & 8 & 9 & 0 & 3 \\ -3 & 5 & 2 & 0 & 1 \\ -5 & -5 & 3 & 0 & 1 \\ 3 & 7 & 3 & 0 & 1 \end{bmatrix}$.

Solution: Using the cofactor expansion along the fourth column gives $\det D = 0$.

Part c) in the last example motivates the following definitions which leads to an important result.

DEFINITION
Upper Triangular
Lower Triangular

An $m \times n$ matrix U is said to be **upper triangular** if $u_{ij} = 0$ whenever $i > j$. An $m \times n$ matrix L is said to be **lower triangular** if $l_{ij} = 0$ whenever $i < j$.

EXAMPLE 6

The matrices $\begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 0 & 2 & 5 \\ 0 & 0 & -1 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, and $\begin{bmatrix} -4 \end{bmatrix}$ are all upper triangular.

The matrices $\begin{bmatrix} 3 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 0 & -2 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ are all lower triangular.

THEOREM 2

If an $n \times n$ matrix A is upper triangular or lower triangular, then

$$\det A = a_{11}a_{22}\cdots a_{nn}$$

Proof: We prove this by induction on n for upper triangular matrices. The proof for lower triangular matrices is similar.

If A is a 2×2 upper triangular matrix, say $A = \begin{bmatrix} a_{11}a_{12} & \\ 0 & a_{22} \end{bmatrix}$, then $\det A = a_{11}a_{22}$ as required. Assume that if B is an $(n-1) \times (n-1)$ upper triangular matrix, then $\det B = b_{11} \cdots b_{(n-1)(n-1)}$. Let A be an $n \times n$ upper triangular matrix. Expanding the determinant along the first column gives

$$\det A = a_{11}(-1)^{1+1}C_{11} + 0 \cdots + 0 = a_{11} \det A(1,1)$$

But, $A(1,1)$ is the $(n-1) \times (n-1)$ upper triangular matrix formed by deleting the first row and first column of A. Thus, $\det A(1,1) = a_{11}a_{22} \cdots (n-1)(n-1)$ by the inductive hypothesis. Thus, $\det A = a_{11}a_{22} \cdots a_{nn}$ as required. $\square$

Determinants and Row Operations

We have seen that it is very difficult to calculate the determinant of a large matrix with very few zeros, but very easy to calculate the determinant of an upper or lower triangular matrix. Thus, it makes sense to look at how applying elementary row operations to a matrix changes the determinant.

THEOREM 3 Let A be an $n \times n$ matrix and let B be the matrix obtained from A by multiplying one row of A by $c \in \mathbb{R}$. Then $\det B = c \det A$.

Proof: We will prove the theorem by induction on n. For $n = 2$, the result is easily verified. Assume the result holds for all $(n-1) \times (n-1)$ matrices with $n \geq 3$. Let B be the matrix obtained from A by multiplying the ℓ-th row of A by c and let $B(i,j)$ denote the $(n-1) \times (n-1)$ submatrix of B obtained by deleting the i-th row and j-th column of B. Then

$$b_{ij} = \begin{cases} a_{ij} & \text{if } i \neq \ell \\ ca_{ij} & \text{if } i = \ell \end{cases}$$

and $B(i,j)$ is the matrix obtained from $A(i,j)$ by multiplying some row of $A(i,j)$ by c. Hence, by the inductive hypothesis, $\det B(i,j) = c \det A(i,j)$. Therefore, using the cofactor expansion along the i-th row, with $i \neq \ell$, gives

$$\begin{aligned}
\det B &= \sum_{k=1}^{n} b_{ik}(-1)^{i+k} \det B(i,k) \\
&= \sum_{k=1}^{n} a_{ik}(-1)^{i+k} [c \det A(i,k)] \\
&= c \sum_{k=1}^{n} a_{ik}(-1)^{i+k} \det A(i,k) \\
&= c \det A
\end{aligned}$$

$\square$

THEOREM 4 Let A be an $n \times n$ matrix and let B be the matrix obtained from A by swapping two rows of A. Then $\det B = -\det A$.

Proof: The proof is left as an exercise.

COROLLARY 5 If an $n \times n$ matrix A has two identical rows, then $\det A = 0$.

Proof: If the i-th and j-th rows of A are identical, then let B be the matrix obtained from A by adding (-1) times the i-th row of A to the j-th row. Then, B has a row of zeroes, so performing a cofactor expansion along that row gives

$$0 = \det B = \det A$$

$\square$

THEOREM 6 Let A be an $n \times n$ matrix and let B be the matrix obtained from A by adding a multiple of one row of A to another row. Then $\det B = \det A$.

Proof: The proof is left as an exercise.

We now see that we can find a determinant quite efficiently by row reducing the matrix to upper triangular form, keeping track of how the row operations change the determinant. Since adding a multiple of one row to another does not change the determinant, it reduces the chance of errors if this is the only row operation used.

EXAMPLE 7 Evaluate the following determinants.

a) $\begin{vmatrix} 1 & 2 & 3 & 1 \\ -1 & -1 & -1 & 2 \\ 1 & 3 & 1 & 1 \\ -2 & -2 & 0 & -1 \end{vmatrix}$.

Solution: Since adding a multiple of one row to another does not change the determinant, we get

$$\begin{aligned}\begin{vmatrix} 1 & 2 & 3 & 1 \\ -1 & -1 & -1 & 2 \\ 1 & 3 & 1 & 1 \\ -2 & -2 & 0 & -1 \end{vmatrix} &= \begin{vmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & -2 & 0 \\ 0 & 2 & 6 & 1 \end{vmatrix} \\ &= \begin{vmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -4 & -3 \\ 0 & 0 & 2 & -5 \end{vmatrix} \\ &= \begin{vmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -4 & -3 \\ 0 & 0 & 0 & -13/2 \end{vmatrix} \\ &= 1(1)(-4)(-13/2) = 26\end{aligned}$$

b) $\begin{vmatrix} 3 & 2 & -1 & 1 \\ 2 & -1 & 0 & -3 \\ 5 & 2 & 3 & -2 \\ 1 & 3 & -1 & 4 \end{vmatrix}$

Solution: Using the row operation $R_4 + R_2$ and then $R_4 - R_1$, we get

$$\begin{aligned}\begin{vmatrix}3 & 2 & -1 & 1\\ 2 & -1 & 0 & -3\\ 5 & 2 & 3 & -2\\ 1 & 3 & -1 & 4\end{vmatrix} &= \begin{vmatrix}3 & 2 & -1 & 1\\ 2 & -1 & 0 & -3\\ 5 & 2 & 3 & -2\\ 3 & 2 & -1 & 1\end{vmatrix}\\ &= \begin{vmatrix}3 & 2 & -1 & 1\\ 2 & -1 & 0 & -3\\ 5 & 2 & 3 & -2\\ 0 & 0 & 0 & 0\end{vmatrix}\\ &= 0\end{aligned}$$

Determinants and Elementary Matrices

Since multiplying a matrix A on the left by an elementary matrix E performs the same elementary row operation on A as the corresponding row operation for E, Theorems 3, 4, and 6 give the following result.

COROLLARY 7 Let A be an $n \times n$ matrix and let E be an $n \times n$ elementary matrix. Then $\det EA = \det E \det A$.

Proof: Observe from Theorems 3, 4, and 6, that performing an elementary row operation multiplies the determinant of the original matrix by a non-zero real number c (in the case of swapping rows or adding a multiple of one row to another we have $c = -1$ or $c = 1$ respectively). Since E is obtained from I by performing a single row operation, we get that $\det E = c$, where c depends on the single row operation. Similarly, since EA is the matrix obtained from A by performing the same row operation, we have that $\det EA = c \det A = \det E \det A$. $\square$

This corollary allows us to prove many other important results.

THEOREM 8 **(addition to the Invertible Matrix Theorem)**

An $n \times n$ matrix A is invertible if and only if $\det A \neq 0$.

Proof: By Corollary 6 there exists a sequence of elementary matrices $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = R$ where R is the reduced row echelon form of A. Then

$$\det R = \det(E_k \cdots E_1 A) = \det(E_k) \cdots \det(E_1) \det A$$

Since the determinant of an elementary matrix is non-zero we get that $\det A \neq 0$ if and only if $\det R \neq 0$. But $\det R \neq 0$ if and only if $R = I$. Hence, $\det A \neq 0$ if and only if A is invertible. $\square$

THEOREM 9

If A and B are $n \times n$ matrices then $\det(AB) = \det A \det B$.

Proof: If $\det A = 0$, then A is not invertible, so $A\vec{y} = \vec{0}$ has infinitely many solutions by the Invertible Matrix Theorem. Thus $A(B\vec{x}) = \vec{0}$ has infinitely many solutions. So $(AB)\vec{x} = \vec{0}$ has infinitely many solutions, hence AB is not invertible which gives $0 = \det(AB) = \det A \det B$.

If $\det A \neq 0$, then A is invertible, so there exists a sequence of elementary matrices such that $A = E_1 \cdots E_k$. Then

$$\begin{aligned}
\det(AB) &= \det(E_1 \cdots E_k B) \\
&= \det(E_1) \cdots \det(E_k) \det B \\
&= \det(E_1 \cdots E_k) \det B \\
&= \det A \det B
\end{aligned}$$

□

COROLLARY 10

If A is an invertible matrix, then $\det A^{-1} = \frac{1}{\det A}$.

Proof: The proof is left as an exercise.

THEOREM 11

Let A be an $n \times n$ matrix. Then $\det A = \det A^T$.

Proof: The proof is left as an exercise.

Since $\det A = \det A^T$ we can use column operations instead of row operations when evaluating a determinant. In particular, adding a multiple of one column to another does not change the determinant, swapping two columns multiplies the determinant by (-1), and multiplying a column by a scalar c multiplies the determinant by c. Furthermore, combining row operations, column operations, and cofactor expansions can make evaluating determinants quite efficient.

EXAMPLE 8

Evaluate the determinant of $D = \begin{bmatrix} 1 & 3 & -1 & 1 \\ -3 & 2 & 1 & 2 \\ 2 & -1 & 1 & 1 \\ 2 & -3 & 2 & -3 \end{bmatrix}$.

Solution: Using the row operation $R_3 + R_1$ gives

$$\det D = \begin{vmatrix} 1 & 3 & -1 & 1 \\ -3 & 2 & 1 & 2 \\ 3 & 2 & 0 & 2 \\ 2 & -3 & 2 & -3 \end{vmatrix}$$

We now use column operation $C_2 - C_4$ to get

$$\det D = \begin{vmatrix} 1 & 2 & -1 & 1 \\ -3 & 0 & 1 & 2 \\ 3 & 0 & 0 & 2 \\ 2 & 0 & 2 & -3 \end{vmatrix}$$

Using the cofactor expansion along the second column gives

$$\det D = 2(-1)^{1+2} \begin{vmatrix} -3 & 1 & 2 \\ 3 & 0 & 2 \\ 2 & 2 & -3 \end{vmatrix}$$

Now $R_3 - 2R_1$ gives

$$\det D = -2 \begin{vmatrix} -3 & 1 & 2 \\ 3 & 0 & 2 \\ 8 & 0 & -7 \end{vmatrix}$$

Hence, by using the cofactor expansion along the second column we get

$$\det D = (-2)(1)(-1)^{1+2} \begin{vmatrix} 3 & 2 \\ 8 & -7 \end{vmatrix} = -74$$

5.4 Determinants and Systems of Equations

The Invertible Matrix Theorem, tell us that for an $n \times n$ matrix A is invertible if and only if the system of equations $A\vec{x} = \vec{b}$ is consistent with a unique solution for all $\vec{b}$, and that these are both equivalent to $\det A \neq 0$. Thus, it is clear that all three of these concepts are closely related. We now look a little further at this relationship.

Inverse by Cofactors

Observe that we can write the inverse of a matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ in the form

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} \\ C_{12} & C_{22} \end{bmatrix}$$

That is, the entries of A^{-1} are the cofactors of A. In particular,

$$\left(A^{-1}\right)_{ij} = \frac{1}{\det A} C_{ji}$$

It is important to take careful notice of the change of order in the subscripts in the line above. At first, this result may seem suprising, but consider the product

$$\begin{aligned} AA^{-1} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} \\ C_{12} & C_{22} \end{bmatrix} \\ &= \frac{1}{\det A} \begin{bmatrix} a_{11}C_{11} + a_{12}C_{12} & a_{11}C_{21} + a_{12}C_{22} \\ a_{21}C_{11} + a_{22}C_{12} & a_{21}C_{21} + a_{22}C_{22} \end{bmatrix} \\ &= \frac{1}{\det A} \begin{bmatrix} \det A & a_{11}(-a_{12}) + a_{12}a_{11} \\ a_{21}a_{22} + a_{22}(-a_{21}) & \det A \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

We now prove that this formula works in the $n \times n$ case. We begin by proving a lemma.

LEMMA 1 Let A be an $n \times n$ matrix with cofactors C_{ij}. Then

$$\sum_{k=1}^{n} (A)_{ik} C_{jk} = 0$$

Proof: Let B be the matrix obtained from A by replacing the j-th row of A by the i-th row of A. That is, $b_{\ell k} = a_{\ell k}$ for all $\ell \neq j$, and $b_{jk} = a_{ik}$. Then B has two identical rows, so $\det B = 0$. Moreover, since the cofactors of the j-th row of B are the same as the cofactors of the j-th row of A we get

$$0 = \det B = \sum_{k=1}^{n} b_{jk} C_{jk} = \sum_{k=1}^{n} a_{ik} C_{jk}$$

$\square$

In words, this lemma says that if we try to do a cofactor expansion of the determinant, but use the coefficients from one row and the cofactors from another row, then we will always get 0.

THEOREM 2 Let A be an invertible $n \times n$ matrix. Then $(A^{-1})_{ij} = \frac{1}{\det A} C_{ji}$.

Proof: Let B be the $n \times n$ matrix defined by $(B)_{ij} = \frac{1}{\det A} C_{ji}$. Then, for $1 \leq i \leq n$ we have that

$$\left(AB\right)_{ii} = \sum_{k=1}^{n} (A)_{ik} (B)_{ki} = \frac{1}{\det A} \sum_{k=1}^{n} (A)_{ik} C_{ik} = 1$$

For $\big(AB\big)_{ij}$ with $i \neq j$, we have

$$\big(AB\big)_{ij} = \sum_{k=1}^{n}(A)_{ik}(B)_{kj} = \frac{1}{\det A}\sum_{k=1}^{n}(A)_{ik}C_{jk} = 0$$

by Lemma 1. Therefore, $AB = I$ so $B = A^{-1}$. $\square$

DEFINITION
Cofactor Matrix

Let A be an $n \times n$ matrix. Then the **cofactor matrix**, $\operatorname{cof} A$, of A is the matrix whose ij-th entry is the ij-th cofactor of A. That is,

$$(\operatorname{cof} A)_{ij} = C_{ij}$$

DEFINITION
Adjugate

Let A be an $n \times n$ matrix. The **adjugate** of A is the matrix defined by

$$\big(\operatorname{adj} A\big)_{ij} = C_{ji}$$

In particular, $\operatorname{adj} A = (\operatorname{cof} A)^T$.

REMARK

By Theorem 2 we have that $A^{-1} = \frac{1}{\det A}\operatorname{adj} A$.

EXAMPLE 1

Determine the adjugate of $A = \begin{bmatrix} 1 & 0 & 2 \\ -2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix}$ and verify that $A(\operatorname{adj} A) = (\det A)I$.

Solution: We find that the cofactor matrix of A is

$$C = \begin{bmatrix} 1 & 2 & -1 \\ -2 & -5 & 1 \\ -2 & -4 & 1 \end{bmatrix}$$

So

$$\operatorname{adj} A = C^T = \begin{bmatrix} 1 & -2 & -2 \\ 2 & -5 & -4 \\ -1 & 1 & 1 \end{bmatrix}$$

Multiplying we find

$$A(\operatorname{adj} A) = \begin{bmatrix} 1 & 0 & 2 \\ -2 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -2 & -2 \\ 2 & -5 & -4 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Observe that

$$\big(A(\operatorname{adj} A)\big)_{ii} = a_{i1}C_{i1} + a_{i2}C_{i2} + a_{i3}C_{i3} = \det A$$

Thus, we see that $\det A = -1$ and $A(\operatorname{adj} A) = (\det A)I$. Moreover, we have that

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A = \begin{bmatrix} -1 & 2 & 2 \\ -2 & 5 & 4 \\ 1 & -1 & -1 \end{bmatrix}$$

Notice that this is not a very efficient way of calculating the inverse. However, it can be useful in some theoretic applications as it gives a formula for the entries of the inverse.

Cramer's Rule

Consider the system of linear equations $A\vec{x} = \vec{b}$ where A is an $n \times n$ matrix. If A is invertible then we know this has a unique solution

$$\vec{x} = A^{-1}\vec{b} = \frac{1}{\det A}(\operatorname{adj} A)\vec{b} = \frac{1}{\det A} \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Then the i-th component of $\vec{x}$ is

$$x_i = \frac{b_1 C_{1i} + \cdots + b_n C_{ni}}{\det A}.$$

Observe that the quantity $b_1 C_{1i} + \cdots + b_n C_{ni}$ is the determinant of the matrix where we have replaced the i-th column of A by $\vec{b}$. We let

$$A_i = b_1 C_{1i} + \cdots + b_n C_{ni} = \begin{vmatrix} a_{11} & \cdots & a_{1(i-1)} & b_1 & a_{1(i+1)} & \cdots & a_{1n} \\ \vdots & & & \vdots & & & \vdots \\ a_{n1} & \cdots & a_{n(i-1)} & b_n & a_{n(i+1)} & \cdots & a_{nn} \end{vmatrix}$$

and then $x_i = \frac{\det A_i}{\det A}$.

THEOREM 3 **(Cramer's Rule)**

If A is an invertible matrix, then the solution $\vec{x}$ of $A\vec{x} = \vec{b}$ is given by

$$x_i = \frac{\det A_i}{\det A}$$

where A_i is the matrix obtained from A by replacing the i-th column of A by $\vec{b}$.

EXAMPLE 2

Solve the system of linear equations using Cramer's Rule.

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 2 \\ -4x_1 - x_2 + 3x_3 &= 6 \\ -2x_1 + x_2 + 2x_3 &= 3 \end{aligned}$$

Solution: We have $A = \begin{bmatrix} 1 & 2 & 1 \\ -4 & -1 & 3 \\ -2 & 1 & 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$. We find that

$$\det A = \begin{vmatrix} 1 & 2 & 1 \\ -4 & -1 & 3 \\ -2 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 7 & 7 \\ 0 & 5 & 4 \end{vmatrix} = -7$$

Therefore, A is invertible and we can apply Cramer's Rule to get

$$x_1 = \frac{\det A_1}{\det A} = -\frac{1}{7}\begin{vmatrix} 2 & 2 & 1 \\ 6 & -1 & 3 \\ 3 & 1 & 2 \end{vmatrix} = -\frac{1}{7}\begin{vmatrix} 0 & 2 & 1 \\ 0 & -1 & 3 \\ -1 & 1 & 2 \end{vmatrix} = \frac{-7}{-7} = 1$$

$$x_2 = \frac{\det A_2}{\det A} = -\frac{1}{7}\begin{vmatrix} 1 & 2 & 1 \\ -4 & 6 & 3 \\ -2 & 3 & 2 \end{vmatrix} = -\frac{1}{7}\begin{vmatrix} 1 & 0 & 1 \\ -4 & 0 & 3 \\ -2 & -1 & 2 \end{vmatrix} = \frac{7}{-7} = -1$$

$$x_3 = \frac{\det A_3}{\det A} = -\frac{1}{7}\begin{vmatrix} 1 & 2 & 2 \\ -4 & -1 & 6 \\ -2 & 1 & 3 \end{vmatrix} = -\frac{1}{7}\begin{vmatrix} 1 & 2 & 2 \\ 0 & 7 & 14 \\ 0 & 5 & 7 \end{vmatrix} = \frac{-21}{-7} = 3$$

So, the solution is $\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$.

EXAMPLE 3

Let $A = \begin{bmatrix} a & 1 & 2 \\ 0 & b & -1 \\ c & 1 & d \end{bmatrix}$. Assuming that $\det A \neq 0$, solve $A\vec{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

Solution: We have $\det A = abd - 2cb + a - c$. Thus,

$$x_1 = \frac{1}{\det A}\begin{vmatrix} 1 & 1 & 2 \\ -1 & b & -1 \\ 1 & 1 & d \end{vmatrix} = \frac{1}{\det A}\begin{vmatrix} 1 & 1 & 2 \\ 0 & b+1 & 1 \\ 0 & 0 & d-2 \end{vmatrix} = \frac{(b+1)(d-2)}{abd - 2cb + a - c}$$

$$x_2 = \frac{1}{\det A}\begin{vmatrix} a & 1 & 2 \\ 0 & -1 & -1 \\ c & 1 & d \end{vmatrix} = \frac{1}{\det A}\begin{vmatrix} a & 0 & 1 \\ 0 & -1 & -1 \\ c & 0 & d-1 \end{vmatrix} = \frac{-ad + a + c}{abd - 2cb + a - c}$$

$$x_3 = \frac{1}{\det A}\begin{vmatrix} a & 1 & 1 \\ 0 & b & -1 \\ c & 1 & 1 \end{vmatrix} = \frac{1}{\det A}\begin{vmatrix} a & 0 & 1 \\ 0 & b+1 & -1 \\ c & 0 & 1 \end{vmatrix} = \frac{(b+1)(a-c)}{abd - 2cb + a - c}$$

That is, the solution is $\vec{x} = \frac{1}{abd-2cb+a-c}\begin{bmatrix}(b+1)(d-2)\\ -ad+a+c\\ (b+1)(a-c)\end{bmatrix}$.

5.5 Area and Volume

Area of a Parallelogram in $\mathbb{R}^2$

Let $\vec{u} = \begin{bmatrix}u_1\\ u_2\end{bmatrix}$ and $\vec{v} = \begin{bmatrix}v_1\\ v_2\end{bmatrix}$ be two non-zero vectors in $\mathbb{R}^2$. We can then form a parallelogram in $\mathbb{R}^2$ with corner points $(0,0)$, (u_1,u_2), (v_1,v_2), and (u_1+v_1,u_2+v_2). This is called the **paralellogram induced by $\vec{u}$ and $\vec{v}$**. See Figure 5.5.1.

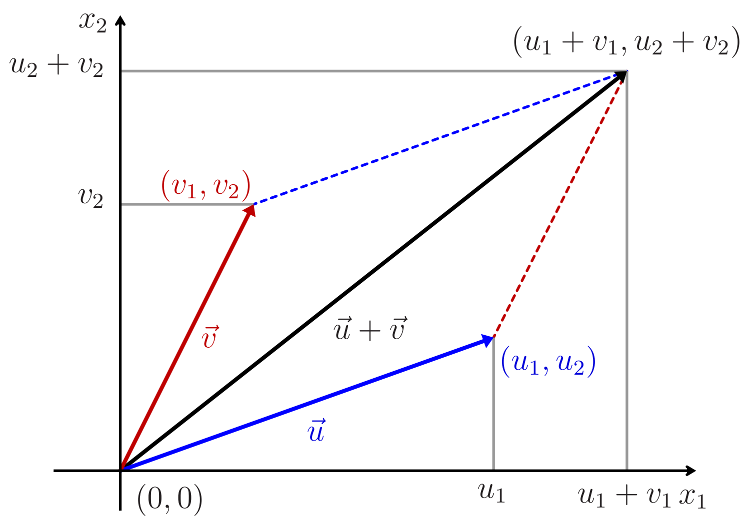

Figure 5.5.1: The parallelogram induced by vectors $\vec{u}$ and $\vec{v}$

We know the area A of a parallelogram is length $\times$ height. As in Figure 5.5.2, we get

$$A = \|\vec{u}\|\|\operatorname{perp}_{\vec{u}}\vec{v}\|$$

Using trigonometry, we see that $\|\operatorname{perp}_{\vec{u}}\vec{v}\| = \|\vec{v}\||\sin\theta|$. Now, recall from Section 1.3

that $\cos\theta = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|}$. Using this gives

$$\begin{aligned}
A &= \|\vec{u}\|\|\vec{v}\||\sin\theta| \\
A^2 &= \|\vec{u}\|^2\|\vec{v}\|^2\sin^2\theta \\
&= \|\vec{u}\|^2\|\vec{v}\|^2(1-\cos^2\theta) \\
&= \|\vec{u}\|^2\|\vec{v}\|^2 - (\vec{u}\cdot\vec{v})^2 \\
&= (u_1^2+u_2^2)(v_1^2+v_2^2) - (u_1v_1+u_2v_2)^2 \\
&= u_1^2v_2^2 + u_2^2v_1^2 - 2(u_1v_2u_2v_1) \\
&= (u_1v_2 - u_2v_1)^2 \\
&= \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix}^2
\end{aligned}$$

Hence, the area of the parallelogram is given by $A = \left|\det\begin{bmatrix}\vec{u} & \vec{v}\end{bmatrix}\right|$.

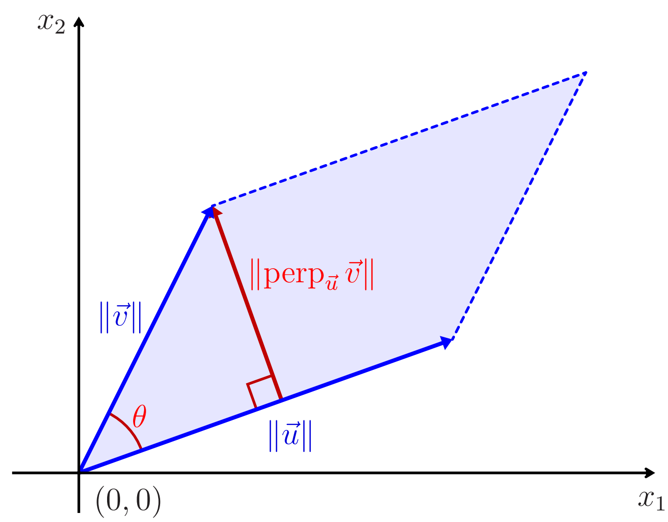

Figure 5.5.2: The area of the parallelogram induced by $\vec{u}$ and $\vec{v}$

EXAMPLE 1

Determine the area of the parallelogram induced by $\vec{u} = \begin{bmatrix}1\\3\end{bmatrix}$ and $\vec{v} = \begin{bmatrix}2\\1\end{bmatrix}$.

Solution: We have

$$A = \left|\det\begin{bmatrix}1 & 2\\3 & 1\end{bmatrix}\right| = |-5| = 5$$

REMARK

Observe that the determinant in the example above gave a negative answer because the angle θ from $\vec{u}$ to $\vec{v}$ was greater than π rads. In particular, if we swap $\vec{u}$ and $\vec{v}$, that is swap the columns of the matrix $\begin{bmatrix}\vec{u} & \vec{v}\end{bmatrix}$, then we will multiply the determinant by -1. In a similar way, we can use the result above to get geometric confirmation of many of the properties of determinants.

Volume of a Parallelepiped

We can repeat what we did above to find the volume of a parallelepiped induced by three vectors $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$ in $\mathbb{R}^3$.

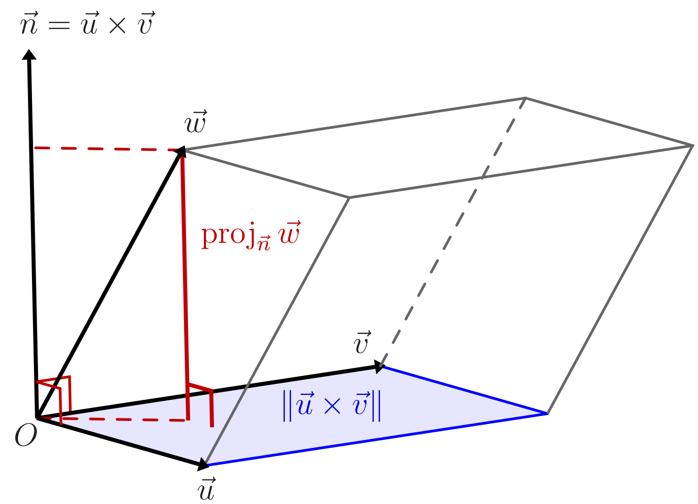

Figure 5.5.3: The volume of a parallelepiped is determined by the area of its base, and by its height

The volume of a parallelepiped is (area of base) $\times$ height. The base is the parallelogram induced by vectors $\vec{u}$ and $\vec{v}$. Thus, using our work above and Theorem 1.3.4, we have that

$$\text{area of base} = \left| \|\vec{u}\| \|\vec{v}\| \sin\theta \right| = \|\vec{u} \times \vec{v}\|$$

Also, the height is the perpendicular of the projection of $\vec{w}$ onto the base. But, this is just the projection of $\vec{w}$ onto the normal vector $\vec{n}$ of the base. Since $\vec{u}$ and $\vec{v}$ lie in the base, we can take $\vec{n} = \vec{u} \times \vec{v}$. Therefore,

$$\text{height} = \|\operatorname{proj}_{\vec{n}} \vec{w}\| = \frac{|\vec{w} \cdot \vec{n}|}{\|\vec{n}\|} = \frac{|\vec{w} \cdot (\vec{u} \times \vec{v})|}{\|\vec{u} \times \vec{v}\|}$$

Hence, the volume V of the parallelepiped is given by

$$\begin{aligned} V &= \frac{|\vec{w} \cdot (\vec{u} \times \vec{v})|}{\|\vec{u} \times \vec{v}\|} \times \|\vec{u} \times \vec{v}\| = |\vec{w} \cdot (\vec{u} \times \vec{v})| \\ &= |w_1(u_2v_3 - u_3v_2) - w_2(u_1v_3 - u_3v_1) + w_3(u_1v_2 - u_2v_1)| \\ &= \left| \det \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix} \right| \end{aligned}$$

EXAMPLE 2

Determine the volume of the parallelepiped determined by $\vec{u} = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} -2 \\ 1 \\ -5 \end{bmatrix}$.

Solution: The volume is

$$V = \left|\det \begin{bmatrix} \vec{u} & \vec{v} & \vec{w} \end{bmatrix}\right| = \left|\det \begin{bmatrix} 1 & 1 & -2 \\ 1 & 3 & 1 \\ 4 & 4 & -5 \end{bmatrix}\right| = \left|\det \begin{bmatrix} 1 & 1 & -2 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}\right| = 6$$

REMARK

Of course, this can be extended to higher dimensions. If $\vec{v}_1, \dots, \vec{v}_n \in \mathbb{R}^n$, then we say that they induce an n-dimensional **parallelotope** (the n-dimensional version of a parallelogram or parallelepiped). The n-**volume** V of the parallelotope is

$$V = \left|\det \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}\right|$$

Chapter 6

Diagonalization

6.1 Matrix of a Linear Mapping and Similar Matrices

Matrix of a Linear Operator

In Chapter 3 we saw how to find the standard matrix of a linear mapping by finding the image of the standard basis vectors under the linear mapping. However, if we look at some particular examples, we see that the standard matrix of a linear mapping is not always nice to work with. We found in Example 3.2.6 that for $\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ the standard matrix of $\mathrm{proj}_{\vec{a}}$ is

$$[\mathrm{proj}_{\vec{a}}] = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}$$

Even though we can use the standard matrix to calculate the projection of a vector $\vec{x}$ onto $\vec{a}$, the matrix itself does not seem to provide us with any information about the action of the projection. This is because we are trying to use the standard basis vectors to determine how a vector is projected onto $\vec{a}$. It would make more sense to instead use a basis for $\mathbb{R}^2$ which contains $\vec{a}$ and a vector orthogonal to $\vec{a}$. Of course, to do this we first need to determine how to find the matrix of a linear operator with respect to a basis other than the standard basis.

Recall that the standard matrix $[L]$ of a linear operator $L : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $L(\vec{x}) = [L]\vec{x}$. To define the matrix of L with respect to any basis $\mathcal{B}$ of $\mathbb{R}^n$, we mimic this formula in $\mathcal{B}$-coordinates instead of standard coordinates. That is, we want

$$[L(\vec{x})]_{\mathcal{B}} = [L]_{\mathcal{B}}[\vec{x}]_{\mathcal{B}}$$

If $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$, then we can write $\vec{x} = b_1\vec{v}_1 + \cdots + b_n\vec{v}_n$. Hence, we have

$$[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \text{ and}$$

$$\begin{aligned}
[L(\vec{x})]_{\mathcal{B}} &= [L(b_1\vec{v}_1 + \cdots + b_n\vec{v}_n)]_{\mathcal{B}} \\
&= [b_1 L(\vec{v}_1) + \cdots + b_n L(\vec{v}_n)]_{\mathcal{B}} \\
&= b_1[L(\vec{v}_1)]_{\mathcal{B}} + \cdots + b_n[L(\vec{v}_n)]_{\mathcal{B}} \\
&= \begin{bmatrix} [L(\vec{v}_1)]_{\mathcal{B}} & \cdots & [L(\vec{v}_n)]_{\mathcal{B}} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}
\end{aligned}$$

Thus, we make the following definition.

DEFINITION

$\mathcal{B}$-Matrix

Let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$ and let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator. Then the **$\mathcal{B}$-matrix** of L is defined to be

$$[L]_{\mathcal{B}} = \begin{bmatrix} [L(\vec{v}_1)]_{\mathcal{B}} & \cdots & [L(\vec{v}_n)]_{\mathcal{B}} \end{bmatrix}$$

EXAMPLE 1

Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear mapping with standard matrix $[L] = \begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix}$ and let $L(\vec{x}) = [L]\vec{x}$. Find the $\mathcal{B}$-matrix of L where $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \right\}$.

Solution: By definition, the columns of the $\mathcal{B}$-matrix are the $\mathcal{B}$-coordinates of the images of the vectors in $\mathcal{B}$ under L. We find that

$$\begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} -6 & -2 & 9 \\ -5 & -1 & 7 \\ -7 & -2 & 10 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ 7 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

Hence,

$$\begin{aligned}
[L]_{\mathcal{B}} &= \begin{bmatrix} [L(1,1,1)]_{\mathcal{B}} & [L(1,0,1)]_{\mathcal{B}} & [L(1,3,2)]_{\mathcal{B}} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}$$

In the example above we had to use the methods of Section 4.3 to find the coordinates of the image vectors with respect to the basis $\mathcal{B}$. However, if we can pick a basis for the linear mapping which is geometrically suited to the mapping, then we can often find the coordinates of the image vectors with respect to the basis by observation.

EXAMPLE 2 Let $\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Determine a geometrically natural basis $\mathcal{B}$ for $\operatorname{proj}_{\vec{a}} : \mathbb{R}^2 \to \mathbb{R}^2$, and then determine $[\operatorname{proj}_{\vec{a}}]_{\mathcal{B}}$.

Solution: As discussed above, it makes sense to include $\vec{a}$ in our geometrically natural basis as well as a vector orthogonal to $\vec{a}$. So, we take the other basis vector to be $\vec{n} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and let $\mathcal{B} = \{\vec{a}, \vec{n}\}$.

By definition, the columns of the $\mathcal{B}$-matrix are the $\mathcal{B}$-coordinates of the images of the vectors in $\mathcal{B}$ under $\operatorname{proj}_{\vec{a}}$. We have

$$\operatorname{proj}_{\vec{a}} \vec{a} = \vec{a} = 1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$
$$\operatorname{proj}_{\vec{a}} \vec{n} = \vec{0} = 0 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

Therefore,

$$[\operatorname{proj}_{\vec{a}}]_{\mathcal{B}} = \begin{bmatrix} [\operatorname{proj}_{\vec{a}} \vec{a}]_{\mathcal{B}} & [\operatorname{proj}_{\vec{a}} \vec{n}]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Let us compare the standard matrix of $\operatorname{proj}_{\vec{a}}$ to its $\mathcal{B}$-matrix found in the example above. First, consider its standard matrix $[\operatorname{proj}_{\vec{a}}] = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}$. Does this matrix give you any useful information about the mapping? It does show us that the projection of $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ onto $\vec{a}$ is $\begin{bmatrix} \frac{1}{5}x_1 + \frac{2}{5}x_2 \\ \frac{2}{5}x_1 + \frac{4}{5}x_2 \end{bmatrix}$. But this does not tell us much about the mapping.

On the other hand, consider its $\mathcal{B}$-matrix $[\operatorname{proj}_{\vec{a}}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. What information does this give us about the mapping? First, this matrix clearly shows the action of the mapping. In particular, for any $\vec{x} = b_1\vec{a} + b_2\vec{n} \in \mathbb{R}^2$ we have

$$[\operatorname{proj}_{\vec{a}} \vec{x}]_{\mathcal{B}} = [\operatorname{proj}_{\vec{a}}]_{\mathcal{B}}[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ 0 \end{bmatrix}$$

That is, the image of $\vec{x}$ under $\operatorname{proj}_{\vec{a}}$ is the amount of $\vec{x}$ in the direction of $\vec{a}$. Also, it is clear from the form of the matrix that the range of the mapping is all scalar multiples of the first vector in $\mathcal{B}$ and the kernel is all scalar multiples of the second vector in $\mathcal{B}$.

The form of the matrix in Example 2 is obviously important. Therefore, we make the following definition.

DEFINITION
Diagonal Matrix

An $n \times n$ matrix D is said to be a **diagonal matrix** if $d_{ij} = 0$ for all $i \neq j$. We denote a diagonal matrix by

$$\operatorname{diag}(d_{11}, d_{22}, \dots, d_{nn})$$

EXAMPLE 3

The diagonal matrix $D = \operatorname{diag}(1, 2, 0, 1)$ is $D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Let's consider another example. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear mapping with standard matrix $[L] = \begin{bmatrix} 1 & 0 & -1 \\ -1/3 & 2/3 & -1/3 \\ -2/3 & -2/3 & 4/3 \end{bmatrix}$. Without doing any further work, what can you determine about L from $[L]$? Not much. However, it can be shown that there exists a basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ for $\mathbb{R}^3$ such that $[L]_\mathcal{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. What can you say about L by looking at $[L]_\mathcal{B}$? We can see that the action of L is to map a vector $\vec{x} = b_1\vec{v}_1 + b_2\vec{v}_2 + b_3\vec{v}_3$ onto $b_1\vec{v}_1 + 2b_2\vec{v}_2 + 0b_3\vec{v}_3$, the range of L is $\operatorname{Span}\{\vec{v}_1, \vec{v}_2\}$, and the kernel of L is $\operatorname{Span}\{\vec{v}_3\}$.

Clearly it is very useful to have a $\mathcal{B}$-matrix of a linear operator $L : \mathbb{R}^n \to \mathbb{R}^n$ in diagonal form. This leads to a couple of questions. How do we determine if there exists a basis $\mathcal{B}$ such that $[L]_\mathcal{B}$ has this form? If there is such a basis, how do we find it? We start answering these questions by looking at our method for calculating the $\mathcal{B}$-matrix a little more closely.

Similar Matrices

Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator with standard matrix $A = [L]$ and let $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$. Then, by definition, we have

$$[L]_\mathcal{B} = \begin{bmatrix} [A\vec{v}_1]_\mathcal{B} & \cdots & [A\vec{v}_n]_\mathcal{B} \end{bmatrix} \tag{6.1}$$

From our work in Section 4.3, we know that the change of coordinates matrix from $\mathcal{B}$-coordinates to $\mathcal{S}$-coordinates is $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$. Hence, we have $[\vec{x}]_\mathcal{B} = P^{-1}\vec{x}$ for any $\vec{x} \in \mathbb{R}^n$. Since $A\vec{v}_i \in \mathbb{R}^n$ for $1 \leq i \leq n$, we can apply this to (6.1) to get

$$\begin{aligned} [L]_\mathcal{B} &= \begin{bmatrix} [A\vec{v}_1]_\mathcal{B} & \cdots & [A\vec{v}_n]_\mathcal{B} \end{bmatrix} \\ &= \begin{bmatrix} P^{-1}A\vec{v}_1 & \cdots & P^{-1}A\vec{v}_n \end{bmatrix} \\ &= P^{-1}A \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} \\ &= P^{-1}AP \\ &= P^{-1}[L]P \end{aligned}$$

REMARK

You can also show that $[L]_{\mathcal{B}} = P^{-1}[L]P$, by showing that $[L]_{\mathcal{B}}[\vec{x}]_{\mathcal{B}} = P^{-1}[L]P[\vec{x}]_{\mathcal{B}}$ for all $\vec{x} \in \mathbb{R}^n$ by simplifying the right-hand side of the equation. This is an excellent exercise.

Since $[L]_{\mathcal{B}}$ and $[L]$ represent the same linear mapping, we would expect that these two matrices have many of the same properties.

THEOREM 1 Let A and B be $n \times n$ matrices such that $P^{-1}AP = B$ for some invertible matrix P. Then

(1) $\operatorname{rank} A = \operatorname{rank} B$
(2) $\det A = \det B$
(3) $\operatorname{tr} A = \operatorname{tr} B$ where $\operatorname{tr} A$ is defined by $\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$ is called the **trace** of a matrix.

Proof: For (1), we multiply both sides on the left of $P^{-1}AP = B$ by P to get $AP = PB$. Then, since P is invertible, it can be written as a product of elementary matrices, say $P = E_1 \cdots E_k$. Hence, by Corollary 5.2.4, we get

$$\operatorname{rank}(PB) = \operatorname{rank}(E_1 \cdots E_k B) = \operatorname{rank}(B)$$

Similarly,

$$\operatorname{rank}(AP) = \operatorname{rank}((AP)^T) = \operatorname{rank}(E_k^T \cdots E_1^T A^T) = \operatorname{rank}(A^T) = \operatorname{rank}(A)$$

Thus, $\operatorname{rank}(A) = \operatorname{rank}(AP) = \operatorname{rank}(PB) = \operatorname{rank}(B)$ as required.

For (2), we have $\det P \neq 0$, so

$$\det(B) = \det(P^{-1}AP) = \det(P^{-1}) \det A \det P = \frac{1}{\det P} \det A \det P = \det A$$

For (3), we first observe that

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} b_{ki} = \sum_{k=1}^{n} \sum_{i=1}^{n} b_{ki} a_{ik} = \operatorname{tr}(BA)$$

Hence,

$$\operatorname{tr}(B) = \operatorname{tr}(P^{-1}AP) = \operatorname{tr}(P^{-1}(AP)) = \operatorname{tr}((AP)P^{-1}) = \operatorname{tr}(APP^{-1}) = \operatorname{tr}(A) \quad \square$$

We see that such matrices are similar in many ways. This motivates the following definition.

DEFINITION
Similar Matrices

Let A and B be $n \times n$ matrices such that $P^{-1}AP = B$ for some invertible matrix P. Then A and B are said to be **similar**.

REMARK

Observe that if $P^{-1}AP = B$, then taking $Q = P^{-1}$ we get that Q is an invertible matrix such that $A = PBP^{-1} = Q^{-1}BQ$. So, the similarity property is symmetric.

We can now rephrase our earlier questions. Given a linear operator $L : \mathbb{R}^n \to \mathbb{R}^n$, is the standard matrix of L similar to a diagonal matrix? If so, how do we find an invertible matrix P such that $P^{-1}[L]P$ is diagonal?

6.2 Eigenvalues and Eigenvectors

Let $A = [L]$ be the standard matrix of a linear operator $L : \mathbb{R}^n \to \mathbb{R}^n$. To determine how to construct an invertible matrix P such that $P^{-1}AP = D$ is diagonal, we work in reverse. That is, we will assume that such a matrix P exists and use this to find what properties P must have.

Let $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$ and let $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 \vec{e}_1 & \cdots & \lambda_n \vec{e}_n \end{bmatrix}$ such that $P^{-1}AP = D$, or alternately $AP = PD$. This gives

$$\begin{aligned} A\begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} &= P\begin{bmatrix} \lambda_1 \vec{e}_1 & \cdots & \lambda_n \vec{e}_n \end{bmatrix} \\ \begin{bmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{bmatrix} &= \begin{bmatrix} \lambda_1 P\vec{e}_1 & \cdots & \lambda_n P\vec{e}_n \end{bmatrix} \\ \begin{bmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{bmatrix} &= \begin{bmatrix} \lambda_1 \vec{v}_1 & \cdots & \lambda \vec{v}_n \end{bmatrix} \end{aligned}$$

Thus, we see that we must have $A\vec{v}_i = \lambda_i \vec{v}_i$ for $1 \leq i \leq n$. Moreover, since P is invertible, the columns of P must be linearly independent. In particular, $\vec{v}_i \neq \vec{0}$ for $1 \leq i \leq n$.

DEFINITION
Eigenvalues
Eigenvectors
Eigenpair

Let A be an $n \times n$ matrix. If there exists a vector $\vec{v} \neq \vec{0}$ such that $A\vec{v} = \lambda\vec{v}$, then λ is called an **eigenvalue** of A and $\vec{v}$ is called an **eigenvector** of A corresponding to λ. The pair $(\lambda, \vec{v})$ is called an **eigenpair**.

If $L : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator with standard matrix $A = [L]$, and if $(\lambda, \vec{v})$ is an eigenpair of A, then observe that we have

$$L(\vec{v}) = A\vec{v} = \lambda\vec{v}$$

Hence, it makes sense to also define eigenvalues and eigenvectors for linear operators.

DEFINITION

Eigenvalues

Eigenvectors

Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator. If there exists a vector $\vec{v} \neq \vec{0}$ such that $L(\vec{v}) = \lambda\vec{v}$, then λ is called an **eigenvalue** of L and $\vec{v}$ is called an **eigenvector** of L corresponding to λ.

EXAMPLE 1

Let $\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{n} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Then, we have

$$\begin{aligned} \operatorname{proj}_{\vec{a}} \vec{a} &= 1\vec{a} \\ \operatorname{proj}_{\vec{a}} \vec{n} &= \vec{0} = 0\vec{n} \end{aligned}$$

Hence, $\vec{a}$ is an eigenvector of $\operatorname{proj}_{\vec{a}}$ with eigenvalue 1, and $\vec{n}$ is an eigenvector of $\operatorname{proj}_{\vec{a}}$ with eigenvalue 0.

Observe that it is easy to determine whether a given vector $\vec{v}$ is an eigenvector of a given matrix A. We just need to determine whether the image of $\vec{v}$ under A is a scalar multiple of $\vec{v}$.

EXAMPLE 2

Determine which of the following vectors are eigenvectors of $A = \begin{bmatrix} 2 & -3 & -1 \\ 1 & -2 & -1 \\ 1 & -3 & 0 \end{bmatrix}$.

Determine the eigenvalue associated with each eigenvector.

a) $\vec{v}_1 = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$

Solution: We have

$$A\vec{v}_1 = \begin{bmatrix} 2 & -3 & -1 \\ 1 & -2 & -1 \\ 1 & -3 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$$

So, $\vec{v}_1$ is an eigenvector with eigenvalue $\lambda_1 = 1$.

b) $\vec{v}_2 = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$

Solution: We have

$$A\vec{v}_2 = \begin{bmatrix} 2 & -3 & -1 \\ 1 & -2 & -1 \\ 1 & -3 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -7 \\ -4 \\ -7 \end{bmatrix}$$

So, $A\vec{v}_2$ is not a scalar multiple of $\vec{v}_2$, so it is not an eigenvector.

c) $\vec{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Solution: We have

$$A\vec{v}_3 = \begin{bmatrix} 2 & -3 & -1 \\ 1 & -2 & -1 \\ 1 & -3 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ -2 \end{bmatrix} = (-2) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

So, $\vec{v}_3$ is an eigenvector with eigenvalue $\lambda_2 = -2$.

However, how do we determine whether a given scalar λ is an eigenvalue of a given matrix A? To do this, we would need to determine whether there exists a vector $\vec{v} \neq \vec{0}$ such that $A\vec{v} = \lambda\vec{v}$. At first glance this equation might look like a system of linear equations, but we observe that the unknown vector $\vec{v}$ is actually on both sides of the equation. To turn this into a system of linear equations, we need to move $\lambda\vec{v}$ to the other side. We get

$$\begin{aligned} A\vec{v} - \lambda\vec{v} &= \vec{0} \\ (A - \lambda I)\vec{v} &= \vec{0} \end{aligned}$$

Hence, we have a homogeneous system of linear equations with coefficient matrix $A - \lambda I$. Since the definition requires that $\vec{v} \neq \vec{0}$, we see that λ is an eigenvalue of A if and only if this system has infinitely many solutions. Thus, since $A - \lambda I$ is $n \times n$, we know by the Invertible Matrix Theorem that we require $\det(A - \lambda I) = 0$. Moreover, we see that to find all the eigenvectors associated with the eigenvalue λ, we just need to determine all non-zero vectors $\vec{v}$ such that $(A - \lambda I)\vec{v} = \vec{0}$.

EXAMPLE 3

Find all of the eigenvalues of $A = \begin{bmatrix} 1 & 6 & 3 \\ 0 & -2 & 0 \\ 3 & 6 & 1 \end{bmatrix}$. Determine all eigenvectors associated with each eigenvalue.

Solution: Our work above shows that all eigenvalues of A must satisfy $\det(A - \lambda I) = 0$. Thus, to find all eigenvalues of A, we solve this equation for λ. We have

$$\begin{aligned} 0 = \det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & 6 & 3 \\ 0 & -2-\lambda & 0 \\ 3 & 6 & 1-\lambda \end{vmatrix} \\ &= (-2-\lambda) \begin{vmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{vmatrix} \\ &= -(\lambda+2)(\lambda^2 - 2\lambda - 8) = -(\lambda+2)(\lambda+2)(\lambda-4) \end{aligned}$$

Hence, the eigenvalues are $\lambda_1 = -2$, $\lambda_2 = -2$, and $\lambda_3 = 4$.

To find the eigenvectors associated with the eigenvalue $\lambda_1 = -2$ we solve the homogeneous system $(A - \lambda_1 I)\vec{v} = \vec{0}$. Row reducing the coefficient matrix gives

$$A - \lambda_1 I = \begin{bmatrix} 3 & 6 & 3 \\ 0 & 0 & 0 \\ 3 & 6 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The solution of the homogeneous system is $\vec{v} = s\begin{bmatrix}-2\\1\\0\end{bmatrix} + t\begin{bmatrix}-1\\0\\1\end{bmatrix}$, $s,t \in \mathbb{R}$. Hence, all eigenvectors for λ_1 are all *non-zero* vectors of the form $\vec{v} = s\begin{bmatrix}-2\\1\\0\end{bmatrix} + t\begin{bmatrix}-1\\0\\1\end{bmatrix}$, $s,t \in \mathbb{R}$. Of course, these are also the eigenvectors for λ_2.

Similarly, for $\lambda_3 = 4$ we row reduce the coefficient matrix of the homogeneous system $(A - \lambda_3 I)\vec{v} = \vec{0}$ to get

$$A - \lambda_3 I = \begin{bmatrix}-3 & 6 & 3\\0 & -6 & 0\\3 & 6 & -3\end{bmatrix} \sim \begin{bmatrix}1 & 0 & -1\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$$

The solution of the homogeneous system is $\vec{v} = t\begin{bmatrix}1\\0\\1\end{bmatrix}$, $t \in \mathbb{R}$. Hence, all eigenvectors for λ_3 are $\vec{v} = t\begin{bmatrix}1\\0\\1\end{bmatrix}$, $t \in \mathbb{R}$ with $t \neq 0$.

Observe that for any $n \times n$ matrix A the equation $\det(A - \lambda I)$ will always be an n-th degree polynomial, and the roots of the polynomial are the eigenvalues of A.

DEFINITION

Characteristic Polynomial

Let A be an $n \times n$ matrix. The **characteristic polynomial** of A is the n-th degree polynomial

$$C(\lambda) = \det(A - \lambda I)$$

THEOREM 1

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if $C(\lambda) = 0$.

Proof: The proof is left as an exercise.

We also see that the set of all eigenvectors associated with an eigenvalue λ is the nullspace of $A - \lambda I$ *not including the zero vector.*

DEFINITION

Eigenspace

Let A be an $n \times n$ matrix with eigenvalue λ. We call the nullspace of $A - \lambda I$ the **eigenspace** of λ. The eigenspace is denoted E_λ.

REMARKS

1. It is important to remember that the set of all eigenvectors for the eigenvalue λ of A is all vectors in E_λ excluding the zero vector.
2. Since the eigenspace is just the nullspace of $A - \lambda I$, it is also a subspace of $\mathbb{R}^n$.

EXAMPLE 4

Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Find all eigenvalues of A and determine a basis for the eigenspace of each eigenvalue.

Solution: To find all of the eigenvalues, we find the roots of the characteristic polynomial. We have

$$0 = C(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 0 & 1 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^3$$

Hence, all three eigenvalues of A are 1. For $\lambda_1 = 1$, we have

$$A - \lambda_1 I = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$.

Observe in the examples above that the dimension of the eigenspace of an eigenvalue is always less than or equal to the number of times the eigenvalue is repeated as a root of the characteristic polynomial. Before we prove this is always the case, we make the following definitions and prove a lemma.

DEFINITION

Algebraic Multiplicity

Geometric Multiplicity

Let A be an $n \times n$ matrix with eigenvalue λ_1. The **algebraic multiplicity** of λ_1, denoted a_{λ_1}, is the number of times that λ_1 is a root of the characteristic polynomial $C(\lambda)$. That is, if $C(\lambda) = (\lambda - \lambda_1)^k C_1(\lambda)$, where $C_1(\lambda_1) \neq 0$, then $a_{\lambda_1} = k$. The **geometric multiplicity** of λ, denoted g_{λ_1}, is the dimension of its eigenspace. So, $g_{\lambda_1} = \dim(E_{\lambda_1})$.

LEMMA 2

Let A and B be similar matrices, then A and B have the same characteristic polynomial, and hence the same eigenvalues.

Proof: If A and B are similar, then there exists an invertible matrix P such that $P^{-1}AP = B$. Hence

$$\begin{aligned}\det(B - \lambda I) &= \det(P^{-1}AP - \lambda P^{-1}P)\\ &= \det(P^{-1}(A - \lambda I)P)\\ &= \det P^{-1}\det(A - \lambda I)\det P\\ &= \det(A - \lambda I)\end{aligned}$$

□

THEOREM 3 Let A be an $n \times n$ matrix with eigenvalue λ_1. Then

$$1 \leq g_{\lambda_1} \leq a_{\lambda_1}$$

Proof: First, by definition of an eigenvalue we have that $g_{\lambda_1} = \dim(E_{\lambda_1}) \geq 1$.

Let $\{\vec{v}_1, \ldots, \vec{v}_k\}$ be a basis for E_{λ_1}. Extend this to a basis $\{\vec{v}_1, \ldots, \vec{v}_k, \vec{w}_{k+1}, \ldots, \vec{w}_n\}$ for $\mathbb{R}^n$ and let $P = \begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_k & \vec{w}_{k+1} & \cdots & \vec{w}_n\end{bmatrix}$. Since the columns of P form a basis for $\mathbb{R}^n$, we have that P is invertible by the Inverse Matrix Theorem. Let $F = \begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_k\end{bmatrix}$ and $G = \begin{bmatrix}\vec{w}_{k+1} \cdots & \vec{w}_n\end{bmatrix}$ so that P can be written as the block matrix $P = \begin{bmatrix}F & G\end{bmatrix}$. Similarly, write P^{-1} as the block matrix $P^{-1} = \begin{bmatrix}H\\ J\end{bmatrix}$ so that

$$I = P^{-1}P = \begin{bmatrix}H\\ J\end{bmatrix}\begin{bmatrix}F & G\end{bmatrix} = \begin{bmatrix}HF & HG\\ JF & JG\end{bmatrix}$$

In particular, this gives that $HF = I_k$ and $JF = O_{n-k,n-k}$, the $(n-k)\times(n-k)$ zero matrix. Define the matrix B by $B = P^{-1}AP$, so that B is similar to A.

$$\begin{aligned}B &= P^{-1}AP\\ &= \begin{bmatrix}H\\ J\end{bmatrix} A\begin{bmatrix}F & G\end{bmatrix}\\ &= \begin{bmatrix}H\\ J\end{bmatrix}\begin{bmatrix}AF & AG\end{bmatrix}\\ &= \begin{bmatrix}H\\ J\end{bmatrix}\begin{bmatrix}\lambda_1 F & AG\end{bmatrix}\\ &= \begin{bmatrix}\lambda_1 HF & HAG\\ \lambda_1 JF & JAG\end{bmatrix}\\ &= \begin{bmatrix}\lambda_1 I & X\\ O_{n-k,n-k} & Y\end{bmatrix}\end{aligned}$$

where X is $k \times k$ and Y is $(n-k)\times k$. Then, by Lemma 2, we have the characteristic polynomial $C(\lambda)$ of A equals the characteristic polynomial of B. In particular,

$$C(\lambda) = \det(B - \lambda I) = \begin{vmatrix}(\lambda_1 - \lambda)I & X\\ O_{n-k,n-k} & Y - \lambda I\end{vmatrix}$$

Expanding the determinant we get

$$C(\lambda) = (-1)^k(\lambda - \lambda_1)^k \det(Y - \lambda I)$$

Since, λ_1 may or may not be a root of $\det(Y - \lambda I)$, we get that $a_{\lambda_1} \geq k = g_{\lambda_1}$. $\square$

EXAMPLE 5

In Example 4 we saw that the matrix A has only one eigenvalue $\lambda_1 = 1$ which is a triple root of the characteristic polynomial. Thus, the algebraic multiplicity of λ_1 is $a_{\lambda_1} = 3$. The geometric multiplicity of λ_1 is $g_{\lambda_1} = \dim E_{\lambda_1} = 2$.

EXAMPLE 6

Let $A = \begin{bmatrix} -1 & 6 & 3 \\ 1 & 0 & -1 \\ -3 & 6 & 5 \end{bmatrix}$. Determine the geometric and algebraic multiplicity of all eigenvalues of A.

Solution: We have

$$\begin{aligned}
0 = C(\lambda) &= \begin{vmatrix} -1-\lambda & 6 & 3 \\ 1 & -\lambda & -1 \\ -3 & 6 & 5-\lambda \end{vmatrix} \\
&= (-1-\lambda)\begin{vmatrix} -\lambda & -1 \\ 6 & 5-\lambda \end{vmatrix} + 6(-1)\begin{vmatrix} 1 & -1 \\ -3 & 5-\lambda \end{vmatrix} + 3\begin{vmatrix} 1 & -\lambda \\ -3 & 6 \end{vmatrix} \\
&= (-1-\lambda)(\lambda^2 - 5\lambda + 6) - 6(-\lambda + 2) + 3(6 - 3\lambda) \\
&= -\lambda^3 + 4\lambda^2 - 4\lambda \\
&= -\lambda(\lambda^2 - 4\lambda + 4) \\
&= -\lambda(\lambda - 2)^2
\end{aligned}$$

Thus, the eigenvalues of A are $\lambda_1 = 0$ and $\lambda_2 = 2$. Since λ_1 is a single root of $C(\lambda)$ we have that $a_{\lambda_1} = 1$. We have $a_{\lambda_2} = 2$ since λ_2 is a double root of $C(\lambda)$.

For $\lambda_1 = 0$, we have

$$A - \lambda_1 I = \begin{bmatrix} -1 & 6 & 3 \\ 1 & 0 & -1 \\ -3 & 6 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1/3 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} 1 \\ -1/3 \\ 1 \end{bmatrix} \right\}$. Hence, $g_{\lambda_1} = \dim E_{\lambda_1} = 1$.

For $\lambda_2 = 2$, we have

$$A - \lambda_2 I = \begin{bmatrix} -3 & 6 & 3 \\ 1 & -2 & -1 \\ -3 & 6 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_2} is $\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$. Hence, $g_{\lambda_2} = \dim E_{\lambda_2} = 2$.

Observe in the example above that we needed to factor a cubic polynomial. For purposes of this course, we typically have two options to make this easier. First, we can try to use row and/or column operations when finding $\det(A-\lambda I)$ so that we do not have to factor a cubic polynomial, or alternately we can use the Rational Roots Theorem and synthetic division to factor the cubic. Of course, in the real world we generally have polynomials of degree much larger than 3 and are not so lucky as to have rational roots. In these cases techniques for approximating roots of polynomials or more specialized methods for approximating eigenvalues are employed. These are beyond the scope of this course; however, you should keep in mind that the problems we do here are not really representative of what you may see in the future.

EXAMPLE 7

Let $A = \begin{bmatrix} -8 & -3 & -1 \\ 9 & 4 & -2 \\ -8 & -8 & -1 \end{bmatrix}$. Determine the geometric and algebraic multiplicity of all eigenvalues of A.

Solution: Since adding a multiple of one column to another or adding a multiple of one row to another does not change the determinant, we get

$$\begin{aligned} 0 = C(\lambda) &= \begin{vmatrix} -8-\lambda & -3 & -1 \\ 9 & 4-\lambda & -2 \\ -8 & -8 & -1-\lambda \end{vmatrix} \\ &= \begin{vmatrix} -5-\lambda & -3 & -1 \\ 5+\lambda & 4-\lambda & -2 \\ 0 & -8 & -1-\lambda \end{vmatrix} \\ &= \begin{vmatrix} -5-\lambda & -3 & -1 \\ 0 & 1-\lambda & -3 \\ 0 & -8 & -1-\lambda \end{vmatrix} \\ &= (-5-\lambda)(\lambda^2-25) \\ &= -(\lambda+5)^2(\lambda-5) \end{aligned}$$

Hence, the eigenvalues of A are $\lambda_1 = -5$ and $\lambda_2 = 5$ with $a_{\lambda_1} = 2$ and $a_{\lambda_2} = 1$.

For $\lambda_1 = -5$, we have

$$A - \lambda_1 I = \begin{bmatrix} -3 & -3 & -1 \\ 9 & 9 & -2 \\ -8 & -8 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$. Hence, $g_{\lambda_1} = \dim E_{\lambda_1} = 1$.

For $\lambda_2 = 5$, we have

$$A - \lambda_2 I = \begin{bmatrix} -13 & -3 & -1 \\ 9 & -1 & -2 \\ -8 & -8 & -6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1/8 \\ 0 & 1 & 7/8 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_2} is $\left\{ \begin{bmatrix} 1/8 \\ -7/8 \\ 1 \end{bmatrix} \right\}$. Hence, $g_{\lambda_2} = \dim E_{\lambda_2} = 1$.

6.3 Diagonalization

DEFINITION

Diagonalizable

An $n \times n$ matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix D. If $P^{-1}AP = D$, then we say that P **diagonalizes** A.

REMARK

In this course, we will restrict ourselves to diagonalizing matrices with real entries over $\mathbb{R}$. That is, if A has a complex eigenvalue, then we will say that A is not diagonalizable over $\mathbb{R}$.

Our work in the last section showed that if an $n \times n$ matrix A is diagonalizable, then the columns of P must be linearly independent eigenvectors of A and the diagonal entries of the diagonal matrix D are the eigenvalues of A corresponding columnwise to the eigenvectors in P. In particular, if $\{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis for $\mathbb{R}^n$ of eigenvectors of A corresponding to eigenvalues $\lambda_1, \dots, \lambda_n$, then taking $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$, we get that

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

We know how to find all of the eigenvalues of A, but how do we find a set of n linearly independent eigenvectors of A?

LEMMA 1

Let A be an $n \times n$ matrix with eigenpairs $(\lambda_1, \vec{v}_1), (\lambda_2, \vec{v}_2), \dots, (\lambda_k, \vec{v}_k)$ where $\lambda_i \neq \lambda_j$ for $i \neq j$, then $\{\vec{v}_1, \dots, \vec{v}_k\}$ is linearly independent.

Proof: We will prove this by induction. If $k = 1$, then the result is trivial, since by definition of an eigenvector $\vec{v}_1 \neq \vec{0}$. Assume that the result is true for some $k \geq 1$. To show $\{\vec{v}_1, \dots, \vec{v}_k, \vec{v}_{k+1}\}$ is linearly independent we consider

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k + c_{k+1}\vec{v}_{k+1} = \vec{0} \tag{6.2}$$

Observe that since $A\vec{v}_i = \lambda_i\vec{v}_i$ we have

$$(A - \lambda_i I)\vec{v}_j = A\vec{v}_j - \lambda_i\vec{v}_j = \lambda_j\vec{v}_j - \lambda_i\vec{v}_j = (\lambda_j - \lambda_i)\vec{v}_j$$

for all $1 \leq i, j \leq k+1$. Thus, multiplying both sides of (6.2) by $A - \lambda_{k+1}I$ gives

$$\begin{aligned}\vec{0} &= c_1(\lambda_1 - \lambda_{k+1})\vec{v}_1 + \cdots + c_k(\lambda_k - \lambda_{k+1})\vec{v}_k + c_{k+1}(\lambda_{k+1} - \lambda_{k+1})\vec{v}_{k+1} \\ &= c_1(\lambda_1 - \lambda_{k+1})\vec{v}_1 + \cdots + c_k(\lambda_k - \lambda_{k+1})\vec{v}_k + \vec{0}\end{aligned}$$

By our induction hypothesis, $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is linearly independent thus all the coefficients must be 0. But, $\lambda_i \neq \lambda_{k+1}$ for all $1 \leq i \leq k$, so we must have $c_1 = \cdots = c_k = 0$. Thus (6.2) becomes

$$\vec{0} + c_{k+1}\vec{v}_{k+1} = \vec{0}.$$

But $\vec{v}_{k+1} \neq \vec{0}$ since it is an eigenvector, hence $c_{k+1} = 0$. Therefore, the set is linearly independent. □

THEOREM 2 Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ and let $\mathcal{B}_i = \{\vec{v}_{i,1}, \ldots, \vec{v}_{i,g_{\lambda_i}}\}$ denote a basis for the eigenspace of λ_i for $1 \leq i \leq k$. Then $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ is a linearly independent set.

Proof: We prove the result by induction on k. If $k = 1$, then $\mathcal{B}_1$ is linearly independent since it is a basis for E_{λ_1}. Assume the result is true for some j. That is, assume that $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_j$ is linearly independent and consider $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_j \cup \mathcal{B}_{j+1}$. For simplicity, we will denote the vectors in $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_j$ by $\vec{w}_1, \ldots, \vec{w}_\ell$. Assume that there exists a non-trivial solution to the equation

$$c_1\vec{w}_1 + \cdots + c_\ell\vec{w}_\ell + d_1\vec{v}_{j+1,1} + \cdots + d_{g_{\lambda_{j+1}}}\vec{v}_{j+1,g_{\lambda_{j+1}}} = \vec{0}$$

Observe that $c_1, \ldots, c_\ell$ cannot be all zero as otherwise this would contradict that the fact that $\mathcal{B}_{j+1}$ is a basis. Similarly, $d_1, \ldots, d_{g_{\lambda_{j+1}}}$ cannot be all zero as this would contradict the inductive hypothesis. Thus, we must have at least one c_i non-zero and at least one d_i non-zero. But then we would have a linear combination of some vectors in $\mathcal{B}_{j+1}$ equaling a linear combination of vectors in $\mathcal{B}_1, \ldots, \mathcal{B}_j$ which would contradict Lemma 1. Hence, the solution solution must be the trivial solution. Therefore, $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_j \cup \mathcal{B}_{j+1}$ is linearly independent. The result now follows by induction. □

The theorem shows us how to form a linearly independent set of eigenvectors of a matrix A. We first find a basis for the eigenspace of every eigenvalue of A. Then the set containing all the vectors in every basis is linearly independent by Theorem 2. However, for the matrix P to be invertible, we actually need n linearly independent eigenvectors of A. The next result gives us the condition required for there to exist n linearly independent eigenvectors.

COROLLARY 3 Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then A is diagonalizable if and only if $g_{\lambda_i} = a_{\lambda_i}$ for $1 \leq i \leq k$.

Proof: The proof is left as an exercise.

If λ is an eigenvalue of A such that $g_\lambda < a_\lambda$, then λ is said to be **deficient**. Observe that since $g_\lambda \geq 1$ for every eigenvalue λ, then any eigenvalue λ such that $a_\lambda = 1$ cannot be deficient. This gives the following result.

COROLLARY 4 Let A be an $n \times n$ matrix with n distinct eigenvalues. Then A is diagonalizable.

Proof: The proof is left as an exercise.

We now have an algorithm for diagonalizing a matrix A or showing that it is not diagonalizable.

ALGORITHM

Let A be an $n \times n$ matrix.

1. Find and factor the characteristic polynomial $C(\lambda) = \det(A - \lambda I)$.
2. Let $\lambda_1, \ldots, \lambda_n$ denote the n-roots of $C(\lambda)$ (repeated according to multiplicity). f any of the roots of $C(\lambda)$ are complex, then A is not diagonalizable over $\mathbb{R}$.
3. Find a basis for the eigenspace of each distinct eigenvalue λ by finding the nullspace of $A - \lambda I$.
4. If $g_\lambda < a_\lambda$ for any eigenvalue λ of A, then A is not diagonalizable. Otherwise, form a basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$ of eigenvectors of A by using Theorem 2. Let $P = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$. Then $P^{-1}AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ where λ_i is an eigenvalue corresponding to the eigenvector $\vec{v}_i$ for $1 \leq i \leq n$.

EXAMPLE 1 Show that $A = \begin{bmatrix} -1 & 6 & 3 \\ 1 & 0 & -1 \\ -3 & 6 & 5 \end{bmatrix}$ is diagonalizable and find the invertible matrix P and diagonal matrix D such that $P^{-1}AP = D$.

Solution: In Example 6.2.6 we found that the eigenvalues of A are $\lambda_1 = 0$ and $\lambda_2 = 2$, and that $a_{\lambda_1} = g_{\lambda_1}$ and $a_{\lambda_2} = g_{\lambda_2}$. Hence, A is diagonalizable by Corollary 3. Moreover, we get the the columns of P are the basis vectors from each of the eigenspaces. That is,

$$P = \begin{bmatrix} 1 & 2 & 1 \\ -1/3 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

The diagonal entries in D are the eigenvalues of A corresponding columnwise to the columns in P. Hence,

$$P^{-1}AP = D = \operatorname{diag}(0, 2, 2) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

REMARK

It is important to remember that the answer is not unique. We can arrange the linearly independent eigenvectors of A as the columns of P in any order. We only have to ensure the entry d_{ii} in D is an eigenvalue of A of the eigenvector in the i-th column of P.

EXAMPLE 2 Show that $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is not diagonalizable.

Solution: We have

$$0 = C(\lambda) = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (\lambda - 1)^2$$

Hence, $\lambda_1 = 1$ is the only eigenvalue and $a_{\lambda_1} = 2$. We then get

$$A - \lambda_1 I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Therefore, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$. Therefore $g_{\lambda_1} = 1 < a_{\lambda_1}$, hence A is not diagonalizable.

EXAMPLE 3 Show that $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is not diagonalizable over $\mathbb{R}$.

Solution: We have

$$0 = C(\lambda) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1$$

So, A does not have any real eigenvalues, and is not diagonalizable over $\mathbb{R}$.

REMARK

In Math 235 we will show that $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is diagonalizable over $\mathbb{C}$.

EXAMPLE 4 Show that $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is diagonalizable and find the invertible matrix P and diagonal matrix D such that $P^{-1}AP = D$.

Solution: We have

$$0 = C(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1)$$

Thus, the eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = -1$. Since A has two distinct eigenvalues, it is diagonalizable by Corollary 3.

For $\lambda_1 = 3$, we have
$$A - \lambda_1 I = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

For $\lambda_2 = -1$, we have
$$A - \lambda_2 I = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_2} is $\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$.

We now let $P = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$ and get

$$D = P^{-1}AP = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix}$$

EXAMPLE 5

Show that $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$ is diagonalizable and find the invertible matrix P and diagonal matrix D such that $P^{-1}AP = D$.

Solution: Using column and row operations on the determinant, we get

$$\begin{aligned}
0 = C(\lambda) &= \begin{vmatrix} 1-\lambda & 2 & 2 \\ 2 & 1-\lambda & 2 \\ 2 & 2 & 1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 1-\lambda & 2 & 0 \\ 2 & 1-\lambda & 1+\lambda \\ 2 & 2 & -1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 1-\lambda & 2 & 0 \\ 4 & 3-\lambda & 0 \\ 2 & 2 & -1-\lambda \end{vmatrix} \\
&= (-1-\lambda) \begin{vmatrix} 1-\lambda & 2 \\ 4 & 3-\lambda \end{vmatrix} \\
&= -(\lambda+1)(\lambda^2 - 4\lambda - 5) \\
&= -(\lambda+1)^2(\lambda-5)
\end{aligned}$$

Thus, the eigenvalues of A are $\lambda_1 = -1$ and $\lambda_2 = 5$ where $a_{\lambda_1} = 2$ and $a_{\lambda_2} = 1$.

For $\lambda_1 = -1$, we have

$$A - \lambda_1 I = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_1} is $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$, so $g_{\lambda_1} = a_{\lambda_1}$.

For $\lambda_2 = 5$, we have

$$A - \lambda_2 I = \begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, a basis for E_{λ_2} is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$, so $g_{\lambda_2} = a_{\lambda_2}$.

Therefore, A is diagonalizable by Corollary 3, with $P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ and get

$$D = P^{-1}AP = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

6.4 Powers of Matrices

We now briefly look at one of the many applications of diagonalization. In particular, we will look at how to use diagonalization to easily compute powers of a diagonalizable matrix.

Let A be an $n \times n$ matrix. Then, we define powers of a matrix in the expected way. That is, for any positive integer k, A^k is defined to be A multiplied by itself k times. Thus, for $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix}$ we obviously have

$$A^2 = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 10 \\ -5 & 14 \end{bmatrix}$$

$$A^3 = A^2 A = \begin{bmatrix} -1 & 10 \\ -5 & 14 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} -11 & 38 \\ -19 & 46 \end{bmatrix}$$

$$A^{1000} = \begin{bmatrix} 2^{1001} - 3^{1000} & -2^{1001} + 2 \cdot 3^{1000} \\ 2^{1000} - 3^{1000} & -2^{1000} + 2 \cdot 3^{1000} \end{bmatrix}$$

Perhaps the value of A^{1000} is not so obvious. Moreover, we certainly did not multiply a thousand matrices A together. Instead, we again use the power of the diagonal form of a matrix.

Observe that multiplying a diagonal matrix D by itself is easy. In particular, we have if $D = \operatorname{diag}(d_1, \ldots, d_n)$, then $D^k = \operatorname{diag}(d_1^k, \ldots, d_n^k)$. The following theorem, shows us how we can use this fact with diagonalization to quickly take powers of a diagonalizable matrix A.

THEOREM 1

Let A be an $n \times n$ matrix. If there exists a matrix P and diagonal matrix D such that $P^{-1}AP = D$, then

$$A^k = PD^kP^{-1}$$

Proof: We prove the result by induction on k. If $k = 1$, then $P^{-1}AP = D$ implies $A = PDP^{-1}$ and so the result holds. Assume the result is true for some k. We then have

$$A^{k+1} = A^kA = (PD^kP^{-1})(PDP^{-1}) = PD^kP^{-1}PDP^{-1} = PD^kIDP^{-1} = PD^{k+1}P^{-1}$$

as required. $\square$

EXAMPLE 1

Let $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix}$. Show that $A^{1000} = \begin{bmatrix} 2^{1001} - 3^{1000} & -2^{1001} + 2 \cdot 3^{1000} \\ 2^{1000} - 3^{1000} & -2^{1000} + 2 \cdot 3^{1000} \end{bmatrix}$.

Solution: We first diagonalize A. We have

$$0 = \det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 2 \\ -1 & 4 - \lambda \end{vmatrix} = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3)$$

Thus, the eigenvalues of A are $\lambda_1 = 2$ and $\lambda_2 = 3$.

For $\lambda_1 = 2$ we have

$$A - \lambda_1 I = \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix} \sim \begin{bmatrix} -1 & 2 \\ 0 & 0 \end{bmatrix}$$

so a basis for E_{λ_1} is $\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$.

For $\lambda_2 = 3$ we have

$$A - \lambda_2 I = \begin{bmatrix} -2 & 2 \\ -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

so a basis for E_{λ_2} is $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

Therefore, $P = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$. We compute $P^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ and so

$$\begin{aligned} A^{1000} &= PD^{1000}P^{-1} \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2^{1000} & 0 \\ 0 & 3^{1000} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2^{1001} & 3^{1000} \\ 2^{1000} & 3^{1000} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\ & \begin{bmatrix} 2^{1001} - 3^{1000} & -2^{1001} + 2 \cdot 3^{1000} \\ 2^{1000} - 3^{1000} & -2^{1000} + 2 \cdot 3^{1000} \end{bmatrix} \end{aligned}$$

EXAMPLE 2 Let $A = \begin{bmatrix} -2 & 2 \\ -3 & 5 \end{bmatrix}$. Calculate A^{200}.

Solution: We have

$$0 = \det(A - \lambda I) = \begin{vmatrix} -2-\lambda & 2 \\ -3 & 5-\lambda \end{vmatrix} = \lambda^2 - 3\lambda - 4 = (\lambda + 1)(\lambda - 4).$$

Thus, the eigenvalues of A are $\lambda_1 = -1$ and $\lambda_2 = 4$.

For $\lambda_1 = -1$ we have $A - \lambda_1 I = \begin{bmatrix} -1 & 2 \\ -3 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix}$. so a basis for E_{λ_1} is $\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$.

For $\lambda_2 = 4$ we have $A - \lambda_2 I = \begin{bmatrix} -6 & 2 \\ -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1/3 \\ 0 & 0 \end{bmatrix}$. so a basis for E_{λ_2} is $\left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\}$.

Therefore, $P = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$, $D = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}$. We compute $P^{-1} = \frac{1}{5} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$ and so

$$\begin{aligned} A^{200} = PD^{200}P^{-1} &= \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 4^{200} \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 6 - 4^{200} & -2 + 2 \cdot 4^{200} \\ 3 - 3 \cdot 4^{200} & -1 + 6 \cdot 4^{200} \end{bmatrix} \end{aligned}$$

Index