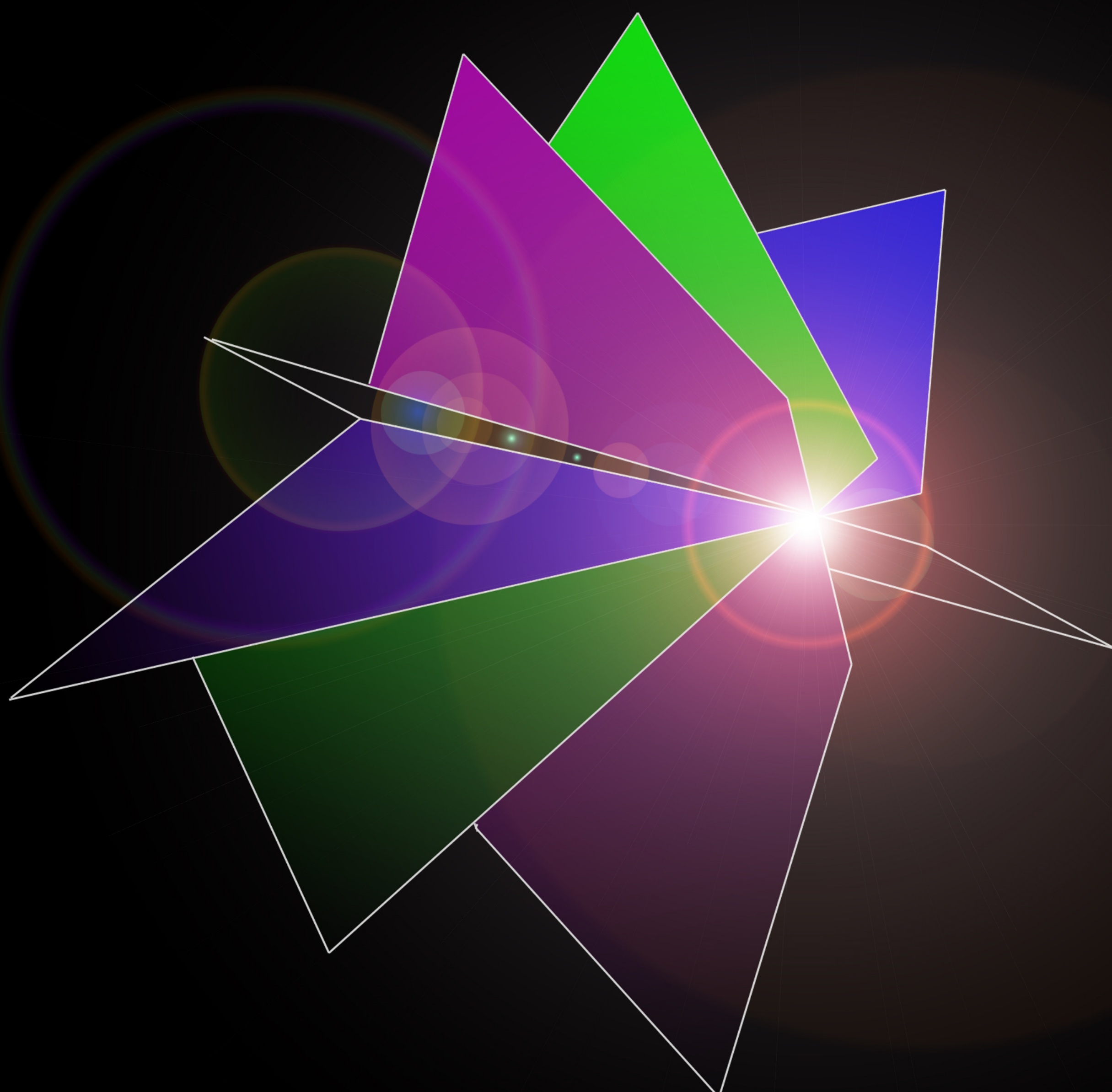
Elementary Linear Algebra

Bruce N. Cooperstein



Elementary Linear Algebra

Bruce N. Cooperstein

ISBN 978-0-9885572-0-8
v.1024120207

Contents

Chapter 1

Linear Equations

1.1. Linear Equations and Their Solution

In this section we review the concepts of a linear equation and linear system. We develop systematic methods for determining when a linear system has a solution and for finding the general solution. Below is a guide to what you find in this section.

Am I Ready for This Material

To be prepared for linear algebra you must have a complete mastery of pre-calculus, some mathematical sophistication, some experience with proofs and the willingness to work hard. More specifically, at the very least you need to be comfortable with algebraic notation, in particular, the use of literals in equations to represent arbitrary numbers as well as the application of subscripts. You should also be able to quickly solve a single linear equation in one variable as well as a system of two linear equations in two variables.

You will also need some mathematical sophistication in order to follow the arguments of the proofs and even attempt some mathematical reasoning of your own.

The most important requirement is that you are willing to perspire, that is, put in the necessary effort. Each section will introduce many new concepts and these have to be understood and mastered before moving on. "Kind of" understanding will not do. Nor will an "intuitive" understanding be sufficient - you will be often asked to follow deductive arguments and even do some proofs yourself. These depend on a deep and genuine understanding of the concepts, and there are lots of them, certainly over fifty major definitions. So, its important to keep up and to periodically return to the definitions and make sure that you truly understand them. With that in mind let me share with you my three "axioms" of time management developed over forty years of teaching (and putting two sons through university):

○ However much time you think you have, you always have less.

○ However much time you think it will take, it will always take more.

○ Something unexpected will come up.

Linear algebra is a beautiful subject and at the same time accessible to first and second year mathematics students. I hope this book conveys the elegance of the subject and is useful to your learning. Good luck.

Before going on to the material of this section you should take the quiz here to see if you have the minimum mathematical skills to succeed in linear algebra.

Quiz

1. Solve the equation

$$3(x-2)+7-2(x+4)=x-11$$

2. Solve the linear equation:

$$4(x-2) - 2(x-4) = 5(x-4) - 4(x-5)$$

3. Solve the linear equation:

$$3(2-x) - 2(3-x) = 4(3-x) - 3(4-x)$$

4. Solve system of two equations in two unknowns:

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

Quiz Solutions

New Concepts

linear equation

solution of linear equation

solution set of a linear equation

equivalent linear equations

standard form of a linear equation

homogeneous linear equation

inhomogeneous linear equation

leading term linear equation in standard form

leading variable of a linear equation in standard form

free variables of a linear equation in standard form

linear system

constants of a linear system

coefficients of a linear system

homogenous linear system

inhomogenous linear system

solution of a linear system

solution set of a linear system

consistent linear system

inconsistent linear system

the trivial solution to a homogeneous linear system

non-trivial solution to a homogeneous linear system

equivalent linear systems

echelon form for a linear system

leading variable

free variable

elementary equation operation

Theory (Why It Works)

Before getting started a brief word about an important convention. This book will be dealing with particular types of equations, linear equations, which we define immediately below. Equations involve "variables". Usually these are real variables which means that we can substitute real numbers for them and the result will be a statement about the equality of two numbers. When there are just a few variables, two or three or possibly four, we will typically use letters at the end of the alphabet, for example x, y for two, x, y, z when there are three variables, and sometimes w, x, y, z when there are four variables. When there are more than a few we will use a single letter but "subscript" the letter with positive whole numbers starting at one, for example, $x_1, x_2, x_3, x_4, x_5, \ldots$ The subscript of such a variable is called its "index." We will also consider our variables to be ordered. The ordering is the natural one for $x_1, x_2, x_3, x_4, \ldots$ while for smaller sets of variables we will use the alphabetical ordering: x, y for two variables, x, y, z for three and w, x, y, z for four.

We now begin our discussion with a definition that is probably familiar from previous mathematics you have studied:

Definition 1.1. A ***linear equation*** is an equation of the form

$$b_1x_1 + b_2x_2 + \cdots + b_nx_n + c = d_1x_1 + d_2x_2 + \cdots + d_nx_n + e \tag{1.1}$$

In Equation (1.1) the variables are $x_1, x_2, \ldots, x_n$. The b_i, d_i stand for (arbitrary but fixed) real numbers and are the coefficients of the variables and c, e are (arbitrary but fixed) real numbers which are the constants of the equation.

An example of the simplest type of **linear equation**, in its most general form, is the following:

$$ax + b = cx + d$$

Example 1.1.1. More concretely we have the equation

$$5x - 11 = 2x + 7 \tag{1.2}$$

When we solve Equation (1.2), that is, find the real numbers we can substitute for x to make both sides the same number, we do the following things:

1. Add $-2x$ to both sides to get

$$3x - 11 = 7 \tag{1.3}$$

Implicit in this operation is that we choose $-2x$ because $2x + (-2x) = 0$.

Secondly, we actually add $-2x$ to $5x - 11$ and $2x + 7$, which respectively yields $(-2x) + (5x - 11)$ and $(-2x) + (2x + 7)$.

We are using the fact that $(-2x)+(5x-11) = (-2x+5x)-11$ and $(-2x)+(2x+7) = (-2x + 2x) + 7$ and also asserting that when equals are added to equals the results are equal.

Now the second expression becomes $(-2 + 2)x + 7$. It then becomes $0x + 7$ and we are using the fact that $0x = 0$ and $0 + 7 = 7$.

Let's return to the equation which has been transformed into $3x - 11 = 7$.

2. We now add 11 to both sides. We choose 11 because $-11 + 11 = 0$. This gives $(3x - 11) + 11 = 7 + 11 = 18$, $3x + (-11 + 11) = 18$, $3x + 0 = 18$ and finally $3x = 18$.

3. Now we divide by 3 or what amounts to the same thing we multiple by $\frac{1}{3}$ to obtain $\frac{1}{3}(3x) = \frac{1}{3} \times 18 = 6$, $[\frac{1}{3} \times 3]x = 6$, and $1x = 6$ whence $x = 6$.

The choice of the $\frac{1}{3}$ was made since it is the number which, when multiplied by 3, yields 1. Note that the last equation has a transparent "solution" (actually, we have not defined this yet), namely, the only value that can be substituted for x in this equation to make it a true statement is the number 6 and this is the solution of the original equation (check this).

To summarize we used the following facts about our system of numbers in **Example** (1.1.1):

(A1) We have an operation called addition which takes two numbers a, b and combines them to obtain the sum $a + b$ which is another number.

(A2) For numbers a, b, c addition satisfies the **associative law** $a+(b+c) = (a+b)+c$

(A3) There is the existence of a **neutral element** for addition also called an additive identity. This means that there is a special number 0 which satisfies $0+a = a$ for every number a.

(A4) For every number there is an **additive inverse**, that is, every number a has an opposite or negative $-a$ such that $a + (-a) = 0$.

(A5) We did not use it, but also addition satisfies the **commutative law** $a + b = b + a$.

(M1) There is also an operation, multiplication, which takes two numbers a, b and combines them to obtain the product ab which is also a number. This operation additionally satisifies:

(M2) Multiplication is **commutative** that is, for any two numbers a, b $ab = ba$.

(M3) Also, multiplication satisfies the **associative law**: for numbers a, b and c we have $a(bc) = (ab)c$.

(M4) There is a **neutral element** for multiplication also called a multiplicative identity, namely the number 1 satisfies $1a = a$ for every number a.

(M5) We also used the fact that every number (excepting 0) has a multiplicative opposite or **inverse** which we denote by $\frac{1}{a}$ or a^{-1} which satisfies $a \times a^{-1} = a^{-1} \times a = 1$.

(M6) Finally, the **distributive law** holds: For any numbers a, b, c we have

$$a(b + c) = ab + ac.$$

The first person to extract these properties and study systems of numbers which satisfy them was **Lagrange**. Such a system was first referred to as a **system of rationality** but today we usually refer to it as a **field**. Examples are: the fractions or rational numbers which we denote by $\mathbb{Q}$, the real numbers, $\mathbb{R}$, and the complex numbers, $\mathbb{C}$. Apart from defining a field, these properties are exceedingly important to us because they are similar to the axioms for a **vector space** which is the central concept in linear algebra.

To remind you the ***rational field*** or simply ***rationals*** consists of all numbers $\frac{m}{n}$ where m, n are integers, $n \neq 0$.

The ***real field*** or ***reals*** are all the possible decimals.

The ***complex field*** consists of all expressions of the form $a + bi$ where a and b are reals and $i = \sqrt{-1}$, that is, satisfies $i^2 = -1$. These are added and multiplied in the following way:

$$(a + bi) + (c + di) = [a + c] + [b + d]i$$

and

$$(a + bi)(c + di) = [ac - bd] + [ad + bc]i.$$

The next simplest type of **linear equation** (and why they are so-called) has the form

$$ax + by + c = dx + ey + f \tag{1.4}$$

By using the operations like those above we can reduce this to one of the forms

$$Ax + By = C, y = A'x + C' \tag{1.5}$$

This should be recognizable from high school mathematics. Because the graph of Equation (1.5) is a **line** we refer to the equation as a linear equation and we extend this definition to any equation in one or more variables in which no power of any variable exceeds 1, that is, there are no $x^2, x^3, \dots$ terms for any variable x and there are no products of distinct variables either (e.g. no terms like xy or x^2y^3.)

As we previously defined (**Definition** (1.1)) the most general form of a linear equation is

$$b_1x_1 + b_2x_2 + \cdots + b_nx_n + c = d_1x_1 + d_2x_2 + \cdots + d_nx_n + e \tag{1.6}$$

Definition 1.2. By a ***solution*** to an equation

$$b_1x_1 + b_2x_2 + \cdots + b_nx_n + c = d_1x_1 + d_2x_2 + \cdots + d_nx_n + e$$

we mean an assignment of actual numbers to the variables

$$x_1 = \gamma_1, x_2 = \gamma_2, \dots, x_n = \gamma_n$$

which makes the two sides equal, that is, makes the proposition a true statement. The ***solution set*** of a linear equation is the collection consisting of all the solutions to the equation.

We will make extensive use of equations which have the same solution set. Because of how frequently we will refer to this relationship we give it a name:

Definition 1.3. Two **linear equations** are said to be ***equivalent*** if they have exactly the same solution sets.

The **general linear equation** is always equivalent to an equation of the form:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1.7}$$

This gives rise to the following definition:

Definition 1.4. A **linear equation** of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ is said to be in ***standard form***.

Example 1.1.2. The equation

$$3x - 2y + z - 5 = 2x - 3y - 2z - 4$$

is **equivalent** to the following equation which in **standard form**.

$$x + y + 3z = 1.$$

Definition 1.5. A **linear equation** like that in **Equation** (1.7) in **standard form** is said to be ***homogeneous*** if $b = 0$ and ***inhomogeneous*** otherwise.

Example 1.1.3. The equation $3x_1 + 4x_2 - x_3 + 7x_4 = 0$ is **homogeneous**. The equation $x_1 - 3x_2 + 5x_3 - 7x_4 = 12$ is **inhomogeneous**.

Definition 1.6. When a **linear equation** is in **standard form** as in **Equation** (1.7) with the variables in their natural order the term with the first nonzero coefficient is called the ***leading term*** and the variable of this term is called the ***leading variable of the equation***. Other variables are **non-leading** variables.

Example 1.1.4. In the **inhomogeneous linear equation**

$$x + y + 3z = 1$$

x is the leading term. Moreover, the triples $(x, y, z) = (1, 6, -2), (0, 1, 0), (3, -5, 1)$ are **solutions**, while (1,1,1) is not a solution.

Example 1.1.5. The equation $0x_1 + 2x_2 - x_3 + 2x_4 = 5$ is an **inhomogeneous linear equation**. The leading term is $2x_2$.

Definition 1.7. In an equation such as those appearing in **Example** (1.1.4) or **Example** (1.1.5) the **non-leading variables** are referred to as ***free variables***. We use this term because these variables can be chosen in any way we wish and once we have done so we can solve for the **leading variable** and get a unique **solution**. Typically, we assign parameters to the free variables and then get an expression for each variable in terms of these parameters. This expression is the **general solution**.

Example 1.1.6. The **leading variable** of the (**standard form**) **linear equation**

$$x + y + 3z = 1$$

is x. We may set $y = s, z = t$ and then $x = 1-s-3t$. Consequently, the **general solution** is $(x, y, z) = (1 - s - 3t, s, t)$ where s and t are free to take on any value.

Example 1.1.7. In the (**standard form**) **linear equation**

$$0x_1 + 2x_2 - x_3 + 2x_4 = 5$$

the **leading variable** is x_2. We can set $x_1 = r, x_3 = s, x_4 = t$ and then $x_2 = \frac{5+s-2t}{2}$ and we have the general solution $(x_1, x_2, x_3, x_4) = (r, \frac{5+s-2t}{2}, s, t)$ where r, s and t can take on any value.

Definition 1.8. A ***system*** of **linear equations** or simply a **linear system** is just a collection of equations like the **standard one**:

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

This system consists of m equations in n variables. The numbers b_i are called the **constants** of the system and the a_{ij} are called the **coefficients** of the system.

If all the $b_i = 0$ then the system is called ***homogeneous*** . Otherwise, if at least one constant is non-zero, it is called ***inhomogeneous***.

Example 1.1.8. The system

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

is an inhomogeneous system of two equations in three variables.

Example 1.1.9. The system

$$\begin{array}{ccccccccc} x_1 & - & 3x_2 & + & 5x_3 & - & 7x_4 & = & 0 \\ 3x_1 & + & 5x_2 & + & 7x_3 & + & 9x_4 & = & 0 \\ 5x_1 & - & 7x_2 & + & 9x_3 & & & = & 0 \end{array}$$

is a homogeneous linear system of three equations in four variables.

Definition 1.9. By a ***solution*** to a **system of linear equations**

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

we mean a sequence of numbers $\gamma_1, \gamma_2, \ldots, \gamma_n$ such that when they are substituted for $x_1, x_2, \ldots, x_n$, all the equations are satisfied, that is, we get equalities.

The set (collection) of all possible solutions is called the ***solution set***. A generic or representative element of the solution set is called a ***general solution***.

Example 1.1.10. Consider the **linear system**

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

Then $(x_1, x_2, x_3) = (-7, 4, 0)$ is a solution, as is $(x_1, x_2, x_3) = (-4, 1, 3)$. However, $(x_1, x_2, x_3) = (0, -2, 7)$ is not a solution.

Definition 1.10. If a **linear system** has a **solution** then it is ***consistent*** otherwise it is ***inconsistent***.

Example 1.1.11. The **linear system** of **Example** (1.1.10) is clearly **consistent**.

The system

$$\begin{array}{ccccc} 2x & + & y & = & 2 \\ 2x & + & y & = & 1 \end{array}$$

is obviously **inconsistent**.

It is less obvious that the following **linear system** is **inconsistent**. We shall shortly develop methods for determining whether a linear system is consistent or not.

$$\begin{array}{ccccccc} x_1 & + & 2x_2 & + & x_3 & = & 4 \\ 2x_1 & + & 3x_2 & + & x_3 & = & 6 \\ x_1 & + & 3x_2 & + & 2x_3 & = & 5 \end{array}$$

To see that the above **linear system** is **inconsistent**, note that

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

If the first two equations are satisfied then we must have

$$x_1 + 3x_2 + 2x_3 = 3 \times 4 - 6 = 6 \neq 5.$$

Definition 1.11. A **homogeneous system of linear equations**

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

always has a **solution**. Namely, set all the variables equal to zero, that is, set $x_1 = 0, x_2 = 0, \ldots, x_n = 0$.

The all -zero solution is called the ***trivial solution***. A solution to a homogeneous system in which some variable is non zero is said to be ***nontrivial***.

Example 1.1.12. The following **homogeneous linear system**

$$\begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & - & 3x_4 & = & 0 \\ x_1 & + & x_2 & - & 3x_3 & + & x_4 & = & 0 \\ x_1 & - & 3x_2 & + & x_3 & + & x_4 & = & 0 \\ -3x_1 & + & x_2 & + & x_3 & + & x_4 & = & 0 \end{array}$$

has the nontrivial solution $(x_1, x_2, x_3, x_4) = (1, 1, 1, 1)$.

Definition 1.12. Two systems of linear equations are said to be ***equivalent*** if their **solution sets** are identical.

Solving a linear system

The way in which we solve a **linear system** such as

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

is to tranform it into an **equivalent linear system** whose **solution set** can be (more) easily determined. The next definition encapsulates just such a type of **linear system**.

Definition 1.13. Consider the following system of linear equations

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

Let l_i be the index of the leading variable of the i^{th} equation. The system is said to have ***echelon form*** if $l_1 < l_2 < \cdots < l_m$, that is, the indices of the leading variables are increasing.

When a linear system is in echelon form, the sequence consisting of those variables which are the leading variable in some equation of the system are the ***leading variables*** of the linear system. The complementary variables (those which are not a leading variable of any of the equations of the linear system) are referred to as the ***free variables*** of the system.

Example 1.1.13. The linear system

$$\begin{array}{rcrcrcr} 2x_1 & + & 3x_2 & + & x_3 & = & 13 \\ & - & 2x_2 & + & 3x_3 & = & 5 \\ & & & & 2x_3 & = & 6 \end{array}$$

is in **echelon form**.

Example 1.1.14. The linear system

$$\begin{array}{rcrcrcrcrl} x_1 & + & 2x_2 & - & 2x_3 & + & x_4 & - & x_5 & = -3 \\ & & x_2 & - & 2x_3 & + & 4x_4 & - & 3x_5 & = 8 \\ & & & & & & 2x_4 & + & x_5 & = 6 \end{array}$$

is in **echelon form**.

Back Substitution

When a linear system is in **echelon form** it can be solved by the method of **back substitution**. In this method each **free variable** of the linear system is set equal to a parameter (that is, it is free to take on any value). Then, beginning with the last equation the leading variable of this equation is solved in terms of the parameters and a constant (possibly zero). This expression is then substituted into the preceding equations and the leading variable in this equation is solved in terms of the parameters. We continue in this way until every leading variable is expressed in terms of the parameters assigned to the free variables (and a constant, possibly zero). The next theorem establishes the theoretic basis for this method.

Theorem 1.1.1. *Let*

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ & & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & \vdots & & \vdots \\ & & & & & & a_{nn}x_n & = & b_n \end{array}$$

be a **linear system** *of n equations in n variables in* **echelon form**. *Assume that $a_{11}, a_{22}, \dots, a_{nn} \neq 0$. Then the system has a unique* **solution**.

Proof. If $n = 1$ then the system simply has the form $ax = b$ with $a \neq 0$ and the only solution is $x = \frac{b}{a}$.

If $n = 2$ then the system looks like

$$\begin{array}{ccccc} a_{11}x & + & a_{12}y & = & b_1 \\ & & a_{22}y & = & b_2 \end{array}.$$

From the second equation we get unique solution for y, namely, $y = \frac{b_2}{a_{22}}$. When this is substituted into the first equation we then get a unique solution for x.

The general proof is obtained by applying the **principle of mathematical induction** to the number of variables (which is equal to the number of equations). The *principle of mathematical induction* says that some proposition (assertion) about **natural numbers** (the counting numbers $1, 2, 3, \dots$) is true for every natural number if we can show it is true for 1 and anytime it is true for a natural number n then we can show that it holds for the next number, $n + 1$.

We have shown that the theorem is true for a single **linear equation** in one variable and a **linear system** of two equations in two variables. Instead of doing the actual proof showing that it holds for $n + 1$ whenever it holds for n, for purposes of illustration we demonstrate how we can go from three variables (and equations) to four variables and equations.

Assume that we know if we have a **linear system** of three equations in three variables in **echelon form** then there is a unique **solution**. Now suppose we are given a system of four equations in four variables which is in echelon form:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & a_{14}x_4 & = & b_1 \\ & & a_{22}x_2 & + & a_{23}x_3 & + & a_{24}x_4 & = & b_2 \\ & & & & a_{33}x_3 & + & a_{34}x_4 & = & b_3 \\ & & & & & & a_{44}x_4 & = & b_4 \end{array}$$

The last equation contains only one variable and so has the unique solution

$x_4 = \frac{b_4}{a_{44}}$. This is the only possible substitution for x_4. Therefore we can substitute this in the previous equations which, after simplification, becomes

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & b_1 - a_{14}\frac{b_4}{a_{44}} \\ & & a_{22}x_2 & + & a_{23}x_3 & = & b_2 - a_{24}\frac{b_4}{a_{44}} \\ & & & & a_{33}x_3 & = & b_3 - a_{34}\frac{b_4}{a_{44}} \end{array}$$

But now we have three equations in three variables in echelon form which, by our assumption, has a unique solution for x_1, x_2, x_3. □

When there are more variables than equations we can assign the **non-leading variables** or **free variables** arbitrarily and once we have done so what remains is **linear system** with equally many variables and equations. By **Theorem** (1.1.1) this system has a unique **solution** which expresses each **leading variable** of the system in term of the free varibles of the system and the constants. Since the free variables can be assigned arbitrarily, in this way we obtain the general solution of the linear system.

As yet unsaid is, for a given **linear system** how do we obtain an **equivalent linear system** in **echelon form**. We discuss this next.

Elementary Equation Operations

Given a **linear system**

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

we obtain an **equivalent linear system** in **echelon form** by sequentially performing certain operations, called ***elementary equation operations***, on the equations of the system. Each elementary equation operation transforms a linear system into an an equivalent linear system. There are three types of elementary equation operations.

Definition 1.14. The elementary equation operations are the following:

1. Multiply an equation by a nonzero scalar (we might do this if every coefficient is divisible by some scalar and then we would multiply by the reciprocal of this scalar. Alternatively, we might do it to make some coefficient equal to one). This is referred to as **scaling**

2. **Exchange** two equations.

3. Add a multiple of one equation to another. We typically do this to eliminate a variable from an equation. This operation is called **elimination**.

It is entirely straightforward that the first two operations will transform a **linear system** into an **equivalent system**: multiplying an equation by a nonzero scalar does not in any way change the solutions of that equation (or any of the other equations). Likewise, if we merely change the order of the equations then we still have the same **solution set**.

That the third operation transforms a **linear system** to an **equivalent linear system** is less obvious. Let's try to see why this works: We illustrate with a simple system of two equations and two unknowns.

Example 1.1.15.

$$\begin{array}{ccccc} 2x & + & 3y & = & 7 \\ 3x & - & y & = & 5 \end{array}$$

Assume that (x_0, y_0) is a common **solution** to both the **linear equations** (for example, $(x_0, y_0) = (2, 1)$). That (x_0, y_0) is a solution to both equations means that when we substitute (plug in) the pair (x_0, y_0) for x and y both the first and second equations are satisifed, that is, we get equalities.

Now if we were to multiply the first equation by -1 then when we plug these numbers in we would get -7. Of course, -7 added to 5 is -2. Consequently, when we plug the solution (x_0, y_0) into the expression

$$-(2x + 3y) + (3x - y)$$

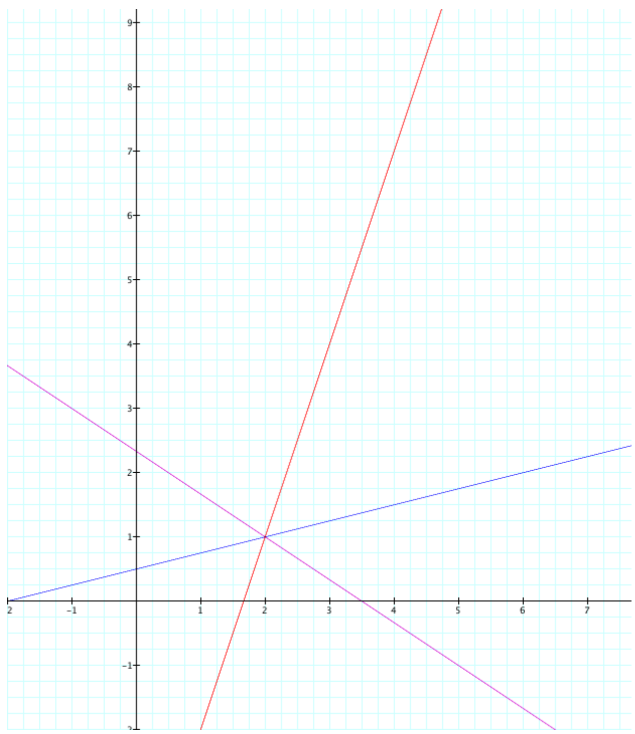
we will get -2. The pair (x_0, y_0) is still a solution to the first equation and therefore it is a solution to the system

$$\begin{array}{ccccc} 2x & + & 3y & = & 7 \\ x & - & 4y & = & -2 \end{array}$$

In **Figure** (1.1.1) we show the graphs of all three equations:

$$\begin{array}{rcrcr} 2x & + & 3y & = & 7 \\ 3x & - & y & = & 5\ . \\ x & - & 4y & = & -2 \end{array}$$

Figure 1.1.1: Intersection of three lines



There is nothing special about -1. We could have multiplied the first equation by -2 or 3 or 6 or anything at all and added to the second and (x_0, y_0) would still be a solution to the resulting system.

This illustrates how every solution to the first system is a solution to the system obtained by application of an elimination operation. However, the process is reversible: We can obtain the first system from

$$\begin{array}{rcrcr} 2x & + & 3y & = & 7 \\ x & - & 4y & = & -2 \end{array}$$

by the same type of operation, specifically, add the first equation to the second. By the same reasoning it follows that every solution to the second system is also a solution to the orginal system. The two systems are equivalent.

We conclude this section with one last theorem which concerns what happens to a homogeneous linear system when we apply elementary equation operations to it.

Theorem 1.1.2. *The application of an* **elementary equation operation** *to a* **homogeneous linear system** *results in a homogeneous system.*

Proof. Suppose the homogeneous linear system is

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned} \tag{1.8}$$

If we exchange the i^{th} and j^{th} equations we exchange their constants which are zero and so all the equations remain homogeneous. If we multiply the i^{th} equation by a scalar c then we multiply its constant by c and the result is $c \times 0 = 0$. Finally if we add c times the i^{th} equation to the j^{th} equation then the resulting constant is c times the constant of the i^{th} equation added to the constant of the j^{th} equation which is $c \times 0 + 0 = 0$. □

In the next section we will demonstrate a procedure using elementary operations which systematically tranforms a system into echelon form (in fact, a particular type of echelon form called reduced echelon form). Then we will also have a method to use the echelon form to find the general solution of the system (or determine if the system is inconsistent).

What You Can Now Do

The main questions answered in this section provide you with the knowledge and methods needed to determine if a **linear system of equations**

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

is **consistent** and, if so, to describe the **solution set** in a compact way. This will be simplified in the next section and then applied in virtually every subsequent section.

In the course of this you will be able to do three kinds of exercises:

1. Given a **linear system** in **echelon form** such as

$$\begin{array}{ccccccc} x & + & 2y & - & z & = & 2 \\ & & 5y & + & z & = & 6 \\ & & & & z & = & 1 \end{array}$$

use **back substitution** to find the (general) **solution** to the system.

2. Use **elementary equation operations** to transform a **linear system** such as

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

into an **equivalent linear system** which is in **echelon form**.

3. Use **elementary equation operations** and **back substitution** to obtain the **general solution** to a **linear system** .

Method (How To Do It)

Method 1.1.1. Use **back substitution** to solve a **linear system** which is in **echelon form**.

Consider the **linear system**

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

Let l_i be the index of the **leading variable** of the i^{th} equation. Assume that the l_i are increasing, that is, $l_1 < l_2 < \cdots < l_m$. If there are equally many variables and equations, $m = n$ then use the last equation to solve for x_n. Substitute this into the preceding equations which gives a linear system of $n-1$ equations in $n-1$ variables. Continue in this way (solve for $x_{n-1,}$ substitute this in the previous equations and so on).

If there are more variables then equations, $n > m$, then determine the **leading variables** and **free variables**. Assign values to the free variables arbitrarily (there are $n-m$ such variables). Once this has been done there are m equations in m variables (the leading variables) and we can proceed by using the method described in the previous paragraph.

Example 1.1.16. The **linear system**

$$\begin{array}{rcrcrcr} 2x_1 & + & 3x_2 & + & x_3 & = & 13 \\ & - & 2x_2 & + & 3x_3 & = & 5 \\ & & & & 2x_3 & = & 6 \end{array}$$

is in **echelon form**.

In this system there are equally many equations as variables. In this case there will be a unique **solution**. We proceed to find it and illustrate how to do **back substitution**

Clearly, in order for the last equation to be satisfied we must set $x_3 = 3$.

Substituting this in the second equation we get

$$-2x_2 + 3(3) = 5$$

which, after doing the arithmetic, becomes

$$-2x_2 = -4$$

and therefore we must have $x_2 = 2$.

In a similar fashion we substitute $x_2 = 2, x_3 = 3$ into the first equation and obtain $x_1 = 2$.

Example 1.1.17. The **linear system**

$$\begin{array}{rcrcrcrcrcr} x_1 & + & 2x_2 & - & 2x_3 & + & x_4 & - & x_5 & = & -3 \\ & & x_2 & - & 2x_3 & + & 4x_4 & - & 3x_5 & = & 8 \\ & & & & & & 2x_4 & + & x_5 & = & 6 \end{array} \tag{1.9}$$

is in **echelon form**. Now there are more variables than there are equations. This will mean that there are infinitely many solutions.

The **leading variable** in the first equation is x_1. The leading variable in the second equation is x_2 and in the third equation the leading variable is x_4. The other (non-leading) variables, x_3 and x_5 are **free variables** because they can be assigned arbitrarily. When this is done we then have three equations in three variables and therefore a unique solution will be obtained for the leading variables x_1, x_2, x_4 in terms of the values assigned to x_3 and x_5 in the same way that a solution was found in **Example** (1.1.16). By setting x_3 and x_5 equal to parameters we obtain the general solution in terms of these parameters.

By way of illustration, we can set $x_3 = x_5 = 0$ and then the equations become

$$\begin{array}{ccccccc} x_1 & + & 2x_2 & + & x_4 & = & -3 \\ & & x_2 & + & 4x_4 & = & 8 \\ & & & & 2x_4 & = & 6 \end{array}$$

Now from the third equation we obtain $x_4 = 3$. Substituting this in the second equation yields

$x_2 + 4 \times 3 = 8$ which when simplified yields $x_2 + 12 = 8$ whence $x_2 = -4$.

Finally substituting $x_2 = -4, x_4 = 3$ into the first equation yields

$x_1 + 2 \times (-4) + 3 = -3$ which, after simplification, becomes $x_1 - 5 = -3$ whence $x_1 = 2$.

Thus the solution we obtain in this case is $(x_1, x_2, x_3, x_4, x_5) = (2, -4, 0, 3, 0)$.

If we substitute $x_3 = 1, x_5 = 2$ the solution is $(x_1, x_2, x_3, x_4, x_5) = (-17, 8, 1, 2, 2)$.

Finding the general solution to a linear system in echelon form.

Example 1.1.18. We now proceed to illustrate how to find the **general solution**. We set the **free variables** equal to parameters: Set $x_3 = s$ and $x_5 = t$. Now the system (1.9) becomes

$$\begin{array}{ccccccccccc} x_1 & + & 2x_2 & - & 2s & + & x_4 & - & t & = & -3 \\ & & x_2 & - & 2s & + & 4x_4 & - & 3t & = & 8 \\ & & & & & & 2x_4 & + & t & = & 6 \end{array}$$

In the last equation we get

$$2x_4 = 6 - t$$

and therefore $x_4 = 3 - \frac{1}{2}t$. This is now substituted into the second equation which becomes

$$x_2 - 2s + 4(3 - \frac{1}{2}t) - 3t = 8$$

After simplification we obtain the equation

$$x_2 - 2s - 5t + 12 = 8$$

and from this we get

$$x_2 = -4 + 2s + 5t$$

Now we substitute $x_2 = -4 + 2s + 5t$ and $x_4 = 3 - \frac{1}{2}t$ into the first equation to obtain

$$x_1 + 2(-4 + 2s + 5t) - 2s + (3 - \frac{1}{2}t) - t = -3$$

which, after simplification, becomes

$$x_1 + 2s + \frac{17}{2}t - 5 = -3$$

from which we conclude that

$$x_1 = 2 - 2s - \frac{17}{2}t.$$

Thus the **general solution** is

$$(x_1, x_2, x_3, x_4, x_5) = (2 - 2s - \frac{17}{2}t, -4 + 2s + 5t, s, 3 - \frac{1}{2}t, t).$$

This means that for any choice of s and t this 5-tuple of numbers is a solution. Check to see that when $x_3 = x_5 = 0$ we obtain the solution (2,-4, 0, 3, 0) and when $x_3 = 1, x_5 = 2$ we get the solution (-17, 8, 1, 2, 2).

Method 1.1.2. Use **elementary equation operations** to put a **linear system** into **echelon form**

There are three **elementary equation operations**. They are the following:

1. Multiply an equation by a nonzero scalar (we might do this if every coefficient is divisible by some scalar and then we would multiply by the reciprocal of this scalar. Alternatively, we might do it to make some coefficient equal to one). This is referred to as **scaling**

2. **Exchange** two equations.

3. Add a multiple of one equation to another. We typically do this to eliminate a variable from an equation. This operation is called **elimination**.

The goal is to use a sequence of these operation to transform the given linear system into a linear system in echelon form.

In the next section we will learn a very precise method for doing this - Gaussian elimination. For now, the idea is to find the first variable that occurs with nonzero coefficient in some equation. This does not have to be x_1. If x_1 has coefficient zero in every equation you might be tempted to drop it. Don't! It has a reason for being there. It will be a **free variable** and we can assign its value arbitrarily though no other variable will involve this parameter.

After finding the first variable that occurs in some equation, use the exchange operation, if necessary, so that this variable occurs in the first equation with a nonzero coefficient. Then use the third operation to eliminate this variable from the equations below it. Then we go on to the remaining equations and play this same game (ignoring the variable we just dealt with since it's coefficient is now zero in all the remaining equations).

It is possible that at some point an equation becomes $0 = 0$. These equations can be dropped (or if you want to to keep the number of equations constant you can put all such equations at the bottom). It is also possible that some equation becomes $0 = 1$. Such an equation would be the last non-zero equation. It means that the system is inconsistent - there are no solutions.

Example 1.1.19. Use **elementary equation operations** to put the following **linear system** into **echelon form**.

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

We add twice the first equation to the second and also add the first equation to the third, after these operations have been performed we obtain the **equivalent linear system**:

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 2 \\ & & 5y & + & z & = & 6 \\ & & 5y & + & 2z & = & 7 \end{array}$$

Now subtract the second equation from the third to obtain the system

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 2 \\ & & 5y & + & z & = & 6 \\ & & & & z & = & 1 \end{array}$$

Example 1.1.20. Put the following **linear system** into **echelon form**:

$$\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ x & + & 2y & + & 3z & = & 2 \\ 2x & + & y & & & = & 1 \end{array}$$

We subtract the first equation from the second and subtract twice the first equation form the third to obtain the **equivalent system**:

$$\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ & & y & + & 2z & = & 1 \\ & & -y & - & 2z & = & -1 \end{array}$$

Now we add the second equation to the which elimates the third equation and yields the system in echelon form:

$$\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ & & y & + & 2z & = & 1 \end{array}$$

or

$$\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ & & y & + & 2z & = & 1 \\ 0x & + & 0y & + & 0z & = & 0 \end{array}$$

Method 1.1.3. Find the **general solution** to a **linear system**

This just combines **Method** (1.1.1) and **Method** (1.1.2):

Use **elementary equation operations** to obtain an **equivalent linear system** in **echelon form**.

If at some point an equation 0 = 1 is obtained, **STOP**, the system is **inconsistent**. Otherwise, once echelon form is obtained use back substitution to find the **general solution**.

Example 1.1.21. Solve the following **linear system**:

$$\begin{array}{rcrcrcr} x & - & 3y & + & 2z & = & 0 \\ 2x & - & 5y & + & 4z & = & 1 \\ x & - & 4y & + & 2z & = & -2 \end{array}$$

We will put this into **echelon form** and then use **back substitution**. We subtract twice the first equation from the second and also subtract the first equation from the third. This yields the following **equivalent linear system**:

$$\begin{array}{rcrcrcr} x & - & 3y & + & 2z & = & 0 \\ & & y & & & = & 1 \\ & & -y & & & = & -2 \end{array}$$

However, now if we add the second equation to the third we obtain the equivalent system

$$\begin{array}{rcrcrcr} x & - & 3y & + & 2z & = & 0 \\ & & y & & & = & 1 \\ & & & & 0 & = & -1 \end{array}$$

which is a contradiction, because 0 cannot equal -1. Consequently, there are no numbers which we can substitute for x, y and z and satisfy all these equations and therefore there is no solution to this system and it is **inconsistent**.

Exercises

In exercises 1-4 use back substitution to find the (general) solution to the linear systems which are given in **echelon form**. See **Method** (1.1.1).

1.

$$\begin{array}{ccccccc} x & + & 2y & - & z & = & 2 \\ & & 5y & + & z & = & 6 \\ & & & & z & = & 1 \end{array}$$

2.

$$\begin{array}{cccccccc} x_1 & -x_2 & +2x_3 & -x_4 & +\ x_5 & = & 6 \\ & 2x_2 & +x_3 & +x_4 & +\ 3x_5 & = & -2 \\ & & x_3 & & +\ x_5 & = & -2 \\ & & & -\ x_4 & -\ x_5 & = & -3 \\ & & & & 2x_5 & = & 4 \end{array}$$

3.

$$\begin{array}{ccccccc} x & + & y & + & z & = & 1 \\ & & y & + & 2z & = & 1 \end{array}$$

4.

$$\begin{array}{ccccccccc} 2w & - & 3x & - & y & + & z & = & 4 \\ & & 2x & + & y & - & z & = & -6 \\ & & & & y & + & 3z & = & -1 \end{array}$$

In exercises 5-8 use **elementary equation operations** to put the system into **echelon form**. See **Method** (1.1.2).

5.

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

6.

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

7.

$$\begin{array}{rcrcrcrcr} w & - & 2x & & & + & 3z & = & 1 \\ 2w & - & 4x & + & y & + & 4z & = & 2 \\ -w & + & 2x & + & y & + & 2z & = & 3 \\ 3w & - & 6x & & & + & 9z & = & 4 \end{array}$$

8.

$$\begin{array}{rcrcrcrcrcr} x_1 & + & x_2 & - & x_3 & + & x_4 & + & 2x_5 & = & 2 \\ 2x_1 & + & 2x_2 & - & x_3 & & & + & 6x_5 & = & 3 \\ 3x_1 & + & 3x_2 & - & x_3 & + & x_4 & + & 9x_5 & = & -1 \end{array}$$

In exercises 9-18 apply **elementary equation operations** to the given linear system to find an **equivalent linear system** in **echelon form**. If the system is consistent then use back substitution to find the general solution. See **Method** (1.1.1) and **Method** (1.1.2).

9.

$$\begin{array}{rcrcrcr} x & + & y & - & 2z & = & 4 \\ 4x & + & 3y & + & z & = & 0 \\ 3x & + & 2y & + & z & = & -2 \end{array}$$

10.

$$\begin{array}{rcrcrcr} 2x & - & 3y & + & z & = & 2 \\ 3x & - & 5y & + & 2z & = & -1 \\ 3x & - & 9y & + & z & = & 7 \end{array}$$

11.

$$\begin{array}{rcrcrcr} x & - & 2y & - & 3z & = & 1 \\ x & - & 3y & - & 8z & = & 2 \\ 2x & - & 2y & + & 3z & = & 4 \end{array}$$

12.

$$
\begin{array}{rcrcrcrcr}
2w & + & x & - & y & + & 3z & = & -3 \\
3w & & & + & y & - & z & = & -2 \\
2w & - & x & + & 3y & - & 5z & = & 1
\end{array}
$$

13.

$$
\begin{array}{rcrcrcrcr}
3w & + & x & & & - & z & = & 2 \\
4w & + & x & - & 2y & + & z & = & 3 \\
w & + & x & + & 4y & - & 5z & = & 0
\end{array}
$$

14.

$$
\begin{array}{rcrcrcrcr}
2w & - & x & + & y & + & 3z & = & 2 \\
5w & - & 3x & + & 2y & + & 6z & = & 5 \\
3w & - & 2x & + & 2y & + & 5z & = & 1 \\
2w & - & x & + & 3y & + & 7z & = & 1
\end{array}
$$

15.

$$
\begin{array}{rcrcrcrcr}
w & + & 2x & + & 3y & - & z & = & 0 \\
2w & + & 3x & - & 2y & + & 2z & = & 0 \\
w & + & 2x & + & 2y & - & 2z & = & 0 \\
 & & x & + & 8y & - & 4z & = & 0
\end{array}
$$

16.

$$
\begin{array}{rcrcrcrcr}
3w & + & 2x & - & 2y & + & z & = & 0 \\
2w & + & 2x & - & 3y & + & z & = & 0 \\
 & & 2x & - & 5y & + & z & = & 0 \\
4w & + & 4x & - & 5y & + & 3z & = & 0
\end{array}
$$

17.

$$
\begin{array}{rcrcrcrcrcr}
2x_1 & + & x_2 & + & x_3 & + & 4x_4 & & & = & 0 \\
3x_1 & + & x_2 & + & 2x_3 & + & 3x_4 & - & 2x_5 & = & 0 \\
4x_1 & + & 2x_2 & + & 2x_3 & + & 7x_4 & - & 3x_5 & = & 0
\end{array}
$$

18.

$$\begin{array}{ccccccccc} w & + & x & + & y & + & z & = & 0 \\ w & - & x & + & 2y & - & 2z & = & 0 \\ w & + & x & + & 4y & + & 4z & = & 0 \\ w & - & x & + & 8y & - & 8z & = & 0 \end{array}$$

In exercises 19-24 indicate whether the statement is true of false.

19. A **homogeneous linear system** of four variables and three equations has infinitely many solutions.

20. A **linear system** of four variables and three equations has infinitely many solutions.

21. A **linear system** with two variables and three equations is **inconsistent**.

22. A **linear system** with three variables and three equations has a unique **solution**.

23. If a **linear system** with four variables and three equations is **consistent** then it has infinitely many solutions.

24. There exists a **linear system** with exactly two **solutions**.

Challenge Exercises (Problems)

1. Write down a **linear system** of three equations in three variables x, y, z such that all the coefficients are non-zero and which has the unique **solution** $(x, y, z) = (-1, 2, -3)$.

2. Write down a **homogeneous linear system** of three distinct equations in three variables that has the **non-trivial solution**$(x, y, z) = (1, -2, 4)$.

3. Write down a **linear system** of two equations in three variables x, y, z which has the **general solution** $(x, y, z) = (3 - 4t, -5 + 2t, t)$.

Quiz Solutions

1. This equation has no solution

2. This equation has the unique solution $x = 0$.

3. Every substituition of a value for x gives a solution since this reduces to the equation 0 = 0.

4. This system has the solution $x = -1, y = 2$.

You should have gotten all of these. If you missed one then you should review your algebra. If you got several wrong you might not be ready for studying Linear Algebra.

1.2. Matrices and Echelon Forms

In this section we introduce one of the most basic objects of linear algebra - a matrix. We show how a **linear system** corresponds to a matrix and translate the **elementary equation operations** of the previous section into corresponding elementary row operations on the matrix of the system. This is used to simplify the procedure for determining the **solution set** of a linear system.

Am I Ready for This Material

Among the concepts that you need to be completely familar with before moving on to the current material are the following:

linear equation

solution of linear equation

solution set of a linear equation

equivalent linear equations

standard form of a linear equation

homogeneous linear equation

inhomogeneous linear equation

leading term linear equation in standard form

leading variable of a linear equation in standard form

free variables of a linear equation in standard form

linear system

constants of a linear system

coefficients of a linear system

homogenous linear system

inhomogenous linear system

solution of a linear system

solution set of a linear system

consistent linear system

inconsistent linear system

the trivial solution to a homogeneous linear system

nontrivial solution to a homogeneous linear system

equivalent linear systems

echelon form for a linear system

leading variable

free variable

elementary equation operation

We also introduced some methods, namely, **Method** (1.1.1) which described how to determine the **solution set** of a **linear system** when it is in **echelon form**, and **Method** (1.1.2), an algorithm for using **elementary equation operations** to put a linear system into echelon form. Before moving on to the new material you should try this quiz to see if you have mastered the methods.

Quiz

1. Find the **solution set and general solution** to the following **linear system**:

$$\begin{array}{rcrcrcr} -3x & + & 5y & + & 4z & = & 5 \\ & & -2y & + & 3z & = & 13 \\ & & & & 2z & = & 6 \end{array}$$

2. Put the following **linear system** into **echelon form**:

$$\begin{array}{rcrcrcrcr} x & + & y & + & z & + & 3w & = & 2 \\ 3x & + & 2y & + & 3z & - & 2w & = & 4 \\ 2x & - & y & + & 3z & + & 2w & = & 3 \end{array}$$

3. Find the **general solution** to the following **linear system**:

$$\begin{array}{rcrcrcrcr} x & + & y & + & 2z & - & 3w & = & -2 \\ 2x & + & 3y & + & 4z & - & 4w & = & 0 \\ x & & & + & 2z & - & 5w & = & -6 \\ 3x & + & 4y & + & 6z & - & 7w & = & -2 \\ & & y & & & + & 2w & = & 4 \end{array}$$

Quiz Solutions

New Concepts

In this section we introduce many new concepts. The most important are:

matrix

entries of a matrix

rows of a matrix

columns of a matrix

zero row of a matrix

non-zero row of a matrix

leading term of a non-zero row of a matrix

coefficient matrix of a linear system

augmented matrix of a linear system

elementary row operation of a matrix

row equivalent matrices

matrix in echelon form

pivot position of a matrix in echelon form

pivot column of a matrix in echelon form

matrix in reduced echelon form

echelon form of a matrix

reduced echelon form of a matrix

pivot positions of a matrix

pivot columns of a matrix

Theory (Why it Works)

In the previous section we began to explore how to solve a **system of linear equations** (also called a **linear system of equations** or simply a **linear system**)

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

through the application of **elementary equation operations** and **back substitution**. Such a linear system can be intimidating with all its variables, plus and minus signs, etc. In point of fact, all the information of the system is encoded by the coefficients of the variables and the constants to the right of the equal signs and as we shall see all the operations can be performed on this collection of numbers. This suggests the following definition:

Definition 1.15. A ***matrix*** is defined to be a rectangular array of numbers. The sequences of numbers which go across the matrix are called ***rows*** and the sequences of numbers that are vertical are called the ***columns*** of the matrix. If there are m rows and n columns then it is said to be an m by n matrix and we write this as $m \times n$.

The numbers which occur in the matrix are called its ***entries*** and the one which is found at the intersection of the i^{th} row and the j^{th} column is called the ij^{th} entry, often written as $(i, j)-$entry.

A ***zero row*** of a matrix is a row all of whose entries are zero and a ***non-zero row*** is a row which has at least one entry which is not zero. The first entry in a non-zero row is called its ***leading entry***.

Consider the following linear system which is in standard form:

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array} \tag{1.10}$$

Everything about the system is “encoded” in the coefficients of the variables and the constants to the right of the equal signs. We can put these numbers into a matrix and this matrix will contain all we need to know about the system. We give a name to this matrix:

Definition 1.16. The matrix

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

consisting of the coefficients of the variables in the linear system (1.10) is called the ***coefficient matrix*** of the system. When the linear system is **inhomogeneous** the matrix

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

obtained by adding a further column consisting of the constants $b_1, b_2, \dots, b_m$ is called the ***augmented matrix*** of the system.

We will generally indicate that a given matrix is augmented by separating the coefficient matrix from the column of constants by a line such as in

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

For a **homogeneous linear system** we do not usually augment since the last column just consists of zeros and as we will see when we perform elementary row operations (to be defined below) on such a matrix the entries in this last column always remain zero.

In the first section we introduced three types of operations on linear systems, called **elementary equation operations**. What was important about these operations is that they transformed a **linear system** into an **equivalent linear system**, that is, a linear system with exactly the same **solution set** as the original system. With each elementary equation operation there is a corresponding operation on the rows of a matrix.

For example, one **elementary equation operation** was referred to as ***exchange*** and involved exchanging two equations, say the i^{th} and j^{th}, in the order of the equations. Corresponding to this is an ***exchange*** elementary row operation which exchanges the i^{th} and j^{th} rows of the matrix.

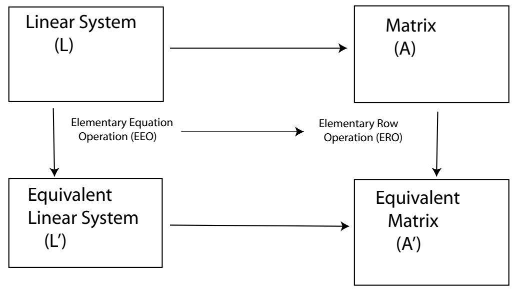


Figure 1.2.1: Elementary operations and the correspondence between linear systems and augmented matrices

Example 1.2.1. If the system of equations is

$$\begin{array}{rcrcrcr} x & + & 3y & - & 4z & = & -5 \\ -2x & + & y & - & 3z & = & 6 \\ 4x & - & 2y & - & z & = & 0 \end{array} \tag{1.11}$$

then the **augmented matrix of the system** is $\begin{pmatrix} 1 & 3 & -4 & | & -5 \\ -2 & 1 & -3 & | & 6 \\ 4 & -2 & -1 & | & 0 \end{pmatrix}$.

If we exchange the second and third equations of (1.11) we obtain the system

$$\begin{array}{rrrcr} x & +3y & -4z & = & -5 \\ 4x & -2y & -z & = & 0 \\ -2x & +y & -3z & = & 6 \end{array} \tag{1.12}$$

The matrix of System (1.12) is $\begin{pmatrix} 1 & 3 & -4 & | & -5 \\ 4 & -2 & -1 & | & 0 \\ -2 & 1 & -3 & | & 6 \end{pmatrix}$

which is obtained from the matrix of by exchanging the second and third rows of this matrix.

In general, suppose L is a **linear system** with **matrix** A and assume we perform an **elementary equation operation** EEO to obtain an **equivalent linear system** L' with matrix A'. Then A' is the matrix obtained from A by applying the elementary row operation corresponding to EEO. This is depicted in the Figure (1.2.1) above.

Recall, in addition to the ***exchange***-type elementary equation operation the others are: ***scaling***, where some equation is multiplied by a non-zero number c; and ***elimination***, whereby a multiple of some equation is added to another equation. We collect in one definition all the corresponding elementary row operations.

Definition 1.17. The corresponding ***elementary row operations*** go by the same names. Thus, we have:

(1) ***Exchange***: Here two rows are interchanged. If the rows are i and j then we denote this by $R_i \leftrightarrow R_j$.

(2) ***Scaling***: Some row is multiplied by a non-zero number. If the i^{th} row is being multiplied by $c \neq 0$ then we write $R_i \to cR_i$.

(3) ***Elimination***: Finally, a multiple of some row is added to another row. If we add c times the i^{th} row to the j^{th} row then we denote this by $R_j \to cR_i + R_j$.

Suppose now that A is the **augmented matrix** for a **linear system** L and A' is the linear system obtained from A by the application of an **elementary row operation** and so corresponds to a system L' which is **equivalent** to L. In this case it is natural to say that the matrices A and A' are **equivalent**. More generally we make the following definition:

Definition 1.18. A matrix A' is said to be ***(row) equivalent*** to the matrix A, denoted by $A \sim A'$, provided A' can be obtained from A by a sequence of elementary row operations.

From what we have already stated it follows that if L and L' are **linear systems** with **corresponding matrices** A and A' which are **row equivalent** then L and L' are **equivalent linear systems**. It will take some work, but we will eventually show that the **converse** of this holds, that is, if L and L' are equivalent linear systems, each consisting of m equations with n variables, with corresponding matrices A and A' then the matrices A and A' are equivalent.

Our immediate goal is to simplify the method for finding the **general solution** of a **linear system**. We previously saw that a linear system can easily be solved by **back substitution** if it is in **echelon form**.

The notion of a **linear system** being in **echelon form** is easily formulated for the corresponding **matrix of the linear system** and inspires the following definition:

Definition 1.19. A matrix A is said to be in ***echelon form*** if the following three conditions hold:

(1) All **zero rows** of A follow all the **non-zero rows** of A;

(2) The **leading entry** in each **non-zero rows** of A is a one; and

(3) The **leading entry** in every row of A subsequent to the first row occurs to the right of the leading entry of the previous row.

If A satisfies these conditions then the leading entry in a non-zero row is referred to as a ***leading one***. The positions where leading ones occur are called the ***pivot positions***. The columns in which a leading one occurs are called ***pivot columns***.

For many theoretic purposes any old echelon form will do, however, for the purpose of solving a **linear system** systematically as well as for other computations a stronger form is desired. This is the subject of the following definition.

Definition 1.20. A matrix is in ***reduced echelon form*** if it is in **row echelon form** and additionally it also satisfies:

(4) All entries above a **leading one** are zero (it follows from condition (3) for a matrix in echelon form that the entries below a leading one are zero).

We have one more definition:

Definition 1.21. If A is a matrix and R is **equivalent** to A and R is in **echelon form** then we will say that R is ***an echelon form of*** A. When R is in reduced echelon form then we will say that R is the ***reduced echelon form of*** A.

A matrix may be **row equivalent** to many different matrices in **echelon form** but, as we shall eventually demonstrate, only one matrix in **reduced echelon form**. This is the reason why, in the above definition we refer to "the" reduced row echelon form of a matrix A.

The practical advantage of having a **reduced echelon form** is that finding the **solution set** of the **corresponding linear system** is exceedingly easy since no **back substitution** is required. We illustrate with an example.

Example 1.2.2. Suppose L is a **linear system** with **augmented matrix**

$$\begin{pmatrix} 1 & 0 & -3 & 0 & 1 & | & 5 \\ 0 & 1 & 2 & 0 & 3 & | & 0 \\ 0 & 0 & 0 & 1 & -4 & | & 3 \end{pmatrix}$$

Then L is the linear system which follows:

$$\begin{array}{ccccccccc} x_1 & & & - & 3x_3 & & & + & x_5 & = 5 \\ & & x_2 & + & 2x_3 & & & + & 3x_5 & = 0 \\ & & & & & & x_4 & - & 4x_5 & = 3 \end{array}$$

The **leading variables** of this linear system are x_1, x_2 and x_4 and the **free variables** are x_3, x_5. We can set $x_3 = s, x_5 = t$ (parameters allowed to assume any value) and then our **general solution** is

$$(x_1, x_2, x_3, x_4, x_5) = (5 + 3s - t, -2s - 3t, s, 3 + 4t, t)$$

As we shall see, it is customary to write a solution in a vertical format (in the next chapter these will be called vectors). So the **general solution** can be expressed as $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 5 + 3s - t \\ -2s - 3t \\ s \\ 3 + 4t \\ t \end{pmatrix}$ where s and t are arbitrary real numbers:

There is a systematic way to apply **elementary row operations** to obtain an **echelon form** (**the reduced echelon form**) of A. The procedure is known as ***Gaussian elimination***. There are five steps to the full Gaussian elimination algorithm. An **echelon form** of a given matrix can be obtained by using just the first four steps; to obtain the **reduced echelon form** the the fifth step may be necessary.

Gaussian Elimination

Let A be an $m \times n$ matrix. Do the following:

(1) Find the first **column**, beginning on the left, which has a non-zero entry. Let this be the j^{th} column.

(2) If the first **row** of the j^{th} column has a zero entry then use an **exchange operation**, exchange the first row with any row which has a non-zero entry in the j^{th} column. In any case, once there is a non-zero entry in the first row of the j^{th} column, use a

scaling operation, if necessary, to make it a one by multiplying the first row by $\frac{1}{a_{1j}}$ and then proceed to step (3).

(3) For each $i \geq 2$ apply an **elimination operation** by adding $-a_{ij}$ times the first row to the i^{th} row to make the $(i, j)-$entry zero.

(4) Next apply the above procedure to the matrix obtained by deleting (covering up) the top row. Repeat this process with all the remaining rows. When this has been completed the matrix will be in **echelon form**.

(5) Starting with the **leading entry** in the last non-zero row work upwards: For each non-zero row introduce zeros above its leading one it by adding appropriate multiples of the row to the rows above it. More specifically, if the $(i, j)-$entry is a leading one, add $-a_{kj}$ times the i^{th} row to the k^{th} row, where $k < i$.

The first four steps of the algorithm are referred to as the ***forward pass***. The fifth step is called the ***backward pass***.

By the first two steps we obtain a non-zero entry at the top of the left-most column that is non-zero and make it a one by rescaling (if necessary). By the second step all the entries below this leading one become zero. After repeating as in step four we obtain a **matrix in echelon form**. Applying step five (the backwards pass) we obtain a **matrix in reduced echelon form**.

Let A be a matrix and assume both E and E' are **echelon forms** of A. We will ultimately see that the (i, j) where the **pivot positions** of E and E' occur are identical. Consequently, the columns in which the pivots positions occur are the same. This allows us to make the following definition for the matrix A:

Definition 1.22. Let A be a matrix and E an **echelon form** of A. The **pivot positions** of E are called the ***pivot positions*** of A and the **pivot columns** of E are called the ***pivot columns*** of A.

Example 1.2.3. The matrix $A = \begin{pmatrix} 1 & 2 & 1 & 1 & 1 \\ 1 & 3 & 0 & 2 & -4 \\ 1 & -1 & 4 & -1 & 1 \\ -3 & -4 & -5 & -2 & 2 \end{pmatrix}$ has **reduced echelon form**

$$E = \begin{pmatrix} 1 & 0 & 3 & 0 & -4 \\ 0 & 1 & -1 & 0 & 10 \\ 0 & 0 & 0 & 1 & -15 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The **pivot positions of** E are (1,1), (2,2) and (3,4). Therefore, (1,1), (2,2) and (3,4) are the pivot positions of A. We also conclude by simple inspection that the **pivot columns** of the matrix E are the $1^{st}, 2^{nd}$, and 4^{th} columns. Consequently, the $1^{st}, 2^{nd}$ and 4^{th} columns are the **pivot columns** of A.

We now state some theorems that give us good information about the **solution set** of a **linear system of equations** from which we can draw some very strong conclusions.

Theorem 1.2.1. *A* **linear system** *is* **consistent** *if and only if the last column of its* **augmented matrix** *is* **not** *a* **pivot column**. *This means that any* **echelon form** *of the augmented matrix* **does not** *have a row of the form* $\begin{pmatrix} 0 & 0 & \dots & 0 & |1 \end{pmatrix}$.

Proof. Let A be the **augmented matrix** of the **linear system**. Assume that the last column of A is a **pivot column**. Then the last non-zero row of the **reduced echelon form** R of A must be

$$\begin{pmatrix} 0 & 0 & \dots & 0 & |1 \end{pmatrix}$$

Then in the **linear system which corresponds to** R (which is **equivalent** to the original linear system) there is an equation of the form

$$0x_1 + 0x_2 + \cdots + 0x_n = 1$$

which has no **solutions**. Therefore our original linear system is **inconsistent**.

On the other hand if no row $\begin{pmatrix} 0 & 0 & \dots & 0 & |1 \end{pmatrix}$ occurs, that is, the last column of A (hence E) is not a **pivot column**, then we can use the procedure for solving a linear system to get one or more **solutions**.

$\square$

We saw in the previous section that when a **linear system** is **consistent**, is in **echelon form**, and there are equally many variables and equations, then every variable is a **leading variable**, there are no **free variables**, and there is a unique **solution**. As a consequence we can state the following theorem.

Theorem 1.2.2. *Assume a* **system of linear equations** *is* **consistent**.

1. The linear system has a unique **solution** *if and only if there are no* **free variables**.

2. The linear system has a unique **solution** *if and only if each column of the* **augmented matrix of the system** *(apart from the last column) is a* **pivot column** *(and the last column is not a pivot column).*

Now a **linear system** can be **inconsistent** (and have no solutions) or be **consistent**. When it is consistent it could have no **free variables** or one or more free variables. By the previous theorem if there are no free variables then the system has a unique **solution**. When there are one or more free variables then the system has infinitely many solutions. These conclusions are summarized in the following theorem:

Theorem 1.2.3. *For a* **system of linear equations** *one and only one of the following hold:*

1. The system is **inconsistent** *(it has no solutions);*
2. The system has a unique **solution***;*
3. The system has infinitely many **solutions***.*

For a **homogeneous system of linear equations** the first possibility cannot occur since there is always the **trivial solution** in which every variable takes the value zero. We summarize the situation for homogeneous linear systems in the following result:

Theorem 1.2.4. Number of solutions of a homogeneous linear system

1. A **homogeneous system of linear equations** *has either only the trivial solution or an infinite number of* **solutions***;*
2. A homogeneous system of linear equations has infinitely many solutions if and only if it has **free variables***;*
3. If a homogeneous system of linear equations has more variables than equations then it has infinitely many solutions.

Remark 1.1. A word of caution: a **homogeneous linear system** of m equations in n variables with $m \geq n$ can have infinitely many **solutions**. The next example illustrates this.

Example 1.2.4. Consider the following homogeneous linear system:

$$\begin{array}{rcrcrcl} 2x & - & 3y & + & z & = & 0 \\ -5x & + & 2y & + & 3z & = & 0 \\ 4x & + & 2y & - & 6z & = & 0 \\ -3x & - & 4y & + & 7z & = & 0 \end{array} \tag{1.13}$$

$(x, y, z) = (t, t, t)$ is a solution for every choice of t.

What You Can Now Do

1. Given a **linear system**

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

write down the **corresponding matrix of the system**.

2. a) Given a matrix A write down the **homogeneous linear system** which has coefficient matrix A.

b) Given a matrix A write down the **inhomogeneous linear system** which has augmented matrix A.

3. Given a matrix A apply **Gaussian Elimination** to obtain an **echelon form** of A and, respectively, the **reduced echelon form** of A.

4. Given a **linear system** determine if it is **consistent or inconsistent**. If a linear system is consistent, find the **general solution** of the system.

Method (How to do it)

Method 1.2.1. Given a **linear system**

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

write down the **corresponding matrix of the system**

Each equation of the **linear system** will correspond to a row of the matrix. If the i^{th} equation is

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$

then the corresponding row of the matrix is $(a_{i1}\ a_{i2}\ \ldots\ a_{in}\ b_i)$. Usually we separate the coefficients of the equation from the constant so it appears as $(a_{i1}\ a_{i2}\ \ldots\ a_{in} \mid b_i)$.

Example 1.2.5. Write down the matrix of the **linear system**

$$\begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & & & + & 2x_4 & = & -2 \\ -2x_1 & & & + & 5x_3 & - & 6x_4 & = & 8 \\ 4x_1 & + & x_2 & - & 2x_3 & - & 3x_4 & = & -4 \end{array}.$$

Remark 1.2. When a variable of the system does not occur in a particular equation of the system its coefficient is zero.

$$\left(\begin{array}{cccc|c} 1 & -3 & 0 & 2 & -2 \\ -2 & 0 & 5 & -6 & 8 \\ 4 & 1 & -2 & -3 & -4 \end{array}\right).$$

Method 1.2.2. a) Given a matrix A write down the **homogeneous linear system** with coefficient matrix A.

b) Given a matrix A write down the **inhomogeneous linear system** with augmented matrix A.

Assign a variable to each of the columns (except the last one in the case the matrix is the **augmented matrix** of an **inhomogeneous linear system**). For each row, multiply each entry by the variable corresponding to the column it is in, add together. Set the resulting linear expression equal to zero in the case the matrix is to be the **coefficient matrix** of a **homogeneous linear system**. If the matrix is to be **augmented matrix** of an **inhomogeneous linear system** then set the linear expression equal to the rightmost entry of the row.

Example 1.2.6. Let $A = \begin{pmatrix} 1 & -1 & 2 & -2 \\ 2 & -3 & 2 & -3 \\ 3 & -1 & 7 & 6 \\ 4 & -3 & 7 & -3 \end{pmatrix}$.

a) Write down the **homogeneous linear system** which has A as its coefficient matrix. The required homogeneous linear system is

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

b) Write down the **inhomogeneous linear system** which has A as it augmented matrix.

Now the linear system is

$$\begin{array}{rcrcrcr} x_1 & - & x_2 & + & 2x_3 & = & -2 \\ 2x_1 & - & 3x_2 & + & 2x_3 & = & -3 \\ 3x_1 & - & x_2 & + & 7x_3 & = & 6 \\ 4x_1 & - & 3x_2 & + & 7x_3 & = & -3 \end{array}$$

Method 1.2.3. Given a matrix A use **Gaussian elimination** to obtain an **echelon form** of A or the **reduced echelon form** of A.

The **method of Gaussian elimination** is fully described **here**.

Example 1.2.7. Use **Gaussian elimination** to obtain the **reduced echelon form of the matrix**.

$$\begin{pmatrix} 0 & 2 & 4 & -6 & 0 & 2 \\ 1 & -1 & -2 & -5 & 2 & -4 \\ 3 & -4 & -2 & -9 & 9 & -7 \\ 2 & -3 & -3 & -7 & -2 & -3 \end{pmatrix}$$

We begin by exchanging rows 1 and 2 ($R_1 \leftrightarrow R_2$)

$$\begin{pmatrix} 1 & -1 & -2 & -5 & 2 & -4 \\ 0 & 2 & 4 & -6 & 0 & 2 \\ 3 & -4 & -2 & -9 & 9 & -7 \\ 2 & -3 & -3 & -7 & -2 & -3 \end{pmatrix}$$

We next add -3 times row one to row three $(R_3 \to (-3)R_1 + R_3)$ and also -2 times row one to row four $(R_4 \to (-2)R_1 + R_4)$. We show the effect of these two operations:

$$\begin{pmatrix} 1 & -1 & -2 & -5 & 2 & -4 \\ 0 & 2 & 4 & -6 & 0 & 2 \\ 0 & -1 & 4 & 6 & 3 & 5 \\ 0 & -1 & 1 & 3 & -6 & 5 \end{pmatrix}$$

We next rescale row 2 by multiplying by $\frac{1}{2}$ $(R_2 \to \frac{1}{2}R_2)$:

$$\begin{pmatrix} 1 & -1 & -2 & -5 & 2 & -4 \\ 0 & 1 & 2 & -3 & 0 & 1 \\ 0 & -1 & 4 & 6 & 3 & 5 \\ 0 & -1 & 1 & 3 & -6 & 5 \end{pmatrix}$$

We now add the second row to the third $(R_3 \to R_2 + R_3)$ and add the second row to the fourth row $(R_4 \to R_2 + R_4)$:

$$\begin{pmatrix} 1 & -1 & -2 & -5 & 2 & -4 \\ 0 & 1 & 2 & -3 & 0 & 1 \\ 0 & 0 & 6 & 3 & 3 & 6 \\ 0 & 0 & 3 & 0 & -6 & 6 \end{pmatrix}$$

Now multiply the third row by $\frac{1}{6}$ $(R_3 \to \frac{1}{6}R_3)$:

$$\begin{pmatrix} 1 & -1 & -2 & -5 & 2 & -4 \\ 0 & 1 & 2 & -3 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 0 & 3 & 0 & -6 & 6 \end{pmatrix}$$

We now add -3 times the third row to the fourth row $(R_4 \to -\frac{1}{2}R_3 + R_4)$:

$$\begin{pmatrix} 1 & -1 & -2 & -5 & 2 & -4 \\ 0 & 1 & 2 & -3 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{15}{2} & 3 \end{pmatrix}$$

Scale the fourth row by multiplying by $-\frac{2}{3}$, $(R_4 \to (-\frac{2}{3})R_4)$:

$$\begin{pmatrix} 1 & -1 & -2 & -5 & 2 & -4 \\ 0 & 1 & 2 & -3 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 0 & 0 & 1 & 5 & -2 \end{pmatrix}$$

We have completed the **forward pass** and this matrix is in **echelon form**. To begin the **backward pass**, add $-\frac{1}{2}$ times the fourth row to the third row $(R_3 \to R_3 + (-\frac{1}{2})R_4)$ to obtain the matrix:

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

Continue with adding 3 times the fourth row to the second, $(R_2 \to R_2 + 3R_4)$; and 5 times the fourth row to the first row, $(R_1 \to R_1 + 5R_4)$:

$$\begin{pmatrix} 1 & -1 & -2 & 0 & 27 & -14 \\ 0 & 1 & 2 & 0 & 15 & -5 \\ 0 & 0 & 1 & 0 & -2 & 2 \\ 0 & 0 & 0 & 1 & 5 & -2 \end{pmatrix}$$

Now add -2 times the third row to the second, $(R_2 \to R_2 + (-2)R_3)$; and add 2 times the third row to the first, $(R_1 \to R_1 + 2R_3)$:

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 23 & -10 \\ 0 & 1 & 0 & 0 & 19 & -9 \\ 0 & 0 & 1 & 0 & -2 & 2 \\ 0 & 0 & 0 & 1 & 5 & -2 \end{pmatrix}$$

Finally, add the second row to the first, $(R_1 \to R_1 + R_2)$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 42 & -19 \\ 0 & 1 & 0 & 0 & 19 & -9 \\ 0 & 0 & 1 & 0 & -2 & 2 \\ 0 & 0 & 0 & 1 & 5 & -2 \end{pmatrix}$$

This matrix is in **reduced echelon form**.

Method 1.2.4. Use matrices to obtain the **general solution** to a **linear system**.

To solve the linear system

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

form the **matrix of the system** (which is **augmented** if the linear system is **inhomogeneous**). Use **Gaussian elimination** to obtain the **reduced echelon form of the matrix**. This is the matrix of an **equivalent linear system**. If the matrix is augmented and the last column of the reduced echelon form is a **pivot column**, **STOP**, the system is **inconsistent** and there are no **solutions**. Otherwise, if the last column is not a pivot column, the system is **consistent**. As we will see this is an easy system to solve.

Write out the corresponding linear system. Then separate the variables into the **leading variables** and the **free variables**. Express each variable in terms of the constants and the free variables. Assign new names to the free variables, that is, make them **parameters** of the solution which can vary arbitrarily. In this way we obtain the general solution of the system.

Example 1.2.8. Find the general solution to the linear system

$$\begin{array}{rcrcrcrcrcl} x_1 & + & 2x_2 & + & x_3 & & & + & 4x_5 & = & 0 \\ 2x_1 & + & 3x_2 & + & x_3 & + & x_4 & + & 7x_5 & = & 1 \\ -2x_1 & + & x_2 & + & 3x_3 & - & 4x_4 & - & 2x_5 & = & -2 \\ x_1 & + & 6x_2 & + & 5x_3 & - & 3x_4 & + & 9x_5 & = & -1 \end{array}.$$

We make the augmented matrix of the system:

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 4 & | & 0 \\ 2 & 3 & 1 & 1 & 7 & | & 1 \\ -2 & 1 & 3 & -4 & -2 & | & -2 \\ 1 & 6 & 5 & -3 & 9 & | & -1 \end{pmatrix}$$

We proceed to use Gaussian elimination to get the reduced echelon form of the matrix. We begin using elimination operations by adding multiples of the first row to the rows below it to make the entries $(i, 1)$ equal to zero for $i > 1$. These operations are $R_2 \to (-2)R_1 + R_2, R_3 \to 2R_1 + R_3, R_4 \to (-1)R_1 + R_4$. The resulting matrix is

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 4 & | & 0 \\ 0 & -1 & -1 & 1 & -1 & | & 1 \\ 0 & 5 & 5 & -4 & 6 & | & -2 \\ 0 & 4 & 4 & -3 & 5 & | & -1 \end{pmatrix}$$

We next scale the second row by the factor -1 , $R_2 \to -R_2$ to obtain

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 4 & | & 0 \\ 0 & 1 & 1 & -1 & 1 & | & -1 \\ 0 & 5 & 5 & -4 & 6 & | & -2 \\ 0 & 4 & 4 & -3 & 5 & | & -1 \end{pmatrix}$$

Now apply elimination by adding multiples of the second row to the third and fourth rows in order to make the (3,2)- and (4,2)-entries zero. The precise operations used are $R_3 \to (-5)R_2 + R_3, R_4 \to (-4)R_2 + R_4$. We get the matrix

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 4 & | & 0 \\ 0 & 1 & 1 & -1 & 1 & | & -1 \\ 0 & 0 & 0 & 1 & 1 & | & 3 \\ 0 & 0 & 0 & 1 & 1 & | & 3 \end{pmatrix}$$

Now we can complete the forward pass by subtracting the third row from the fourth row ($R_4 \to (-1)R_3 + R_4$). This yields the matrix

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

We begin the **backward pass** by adding the third row to the second row ($R_2 \to R_2 + R_3$), thus obtaining the matrix

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

Finally, we add -2 times the second row to the first ($R_1 \to R_1 + (-2)R_2$) and we have a matrix in **reduced echelon form**.

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

The last column is not a **pivot column** so the system is **consistent**. We write out the corresponding system which is **equivalent** to the original system:

$$\begin{array}{ccccccccccc} x_1 & & & - & x_3 & & & & & = & -4 \\ & & x_2 & + & x_3 & & & + & 2x_5 & = & 2 \\ & & & & & & x_4 & + & x_5 & = & 3 \end{array}$$

The **leading variables** are x_1, x_2, x_4 and the **free variables** are x_3, x_5. We set $x_3 = s, x_5 = t$ and substitute:

$$\begin{array}{ccccccccccc} x_1 & & & - & s & & & & & = & -4 \\ & & x_2 & + & s & & & + & 2t & = & 2 \\ & & & & & & x_4 & + & t & = & 3 \end{array}$$

We solve for each of the variables in terms of s and t :

$$\begin{array}{ccc} x_1 & = & -4 + s \\ x_2 & = & 2 - s - 2t \\ x_3 & = & s \\ x_4 & = & 3 - t \\ x_5 & = & t \end{array}$$

It will be convenient to write this as a single column:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -4+s \\ 2-s-2t \\ s \\ 3-t \\ t \end{pmatrix}.$$

Example 1.2.9. Find the general **solution** to the **linear system**

$$\begin{array}{rcrcrcrcr} x_1 & + & 2x_2 & + & x_3 & - & 2x_4 & = & -1 \\ 2x_2 & + & 3x_2 & + & 3x_3 & + & 3x_4 & = & 10 \\ -x_1 & - & 4x_2 & + & x_3 & - & 4x_4 & = & -15 \\ -2x_1 & - & x_2 & - & 5x_3 & + & 3x_4 & = & 5 \end{array}$$

We form the **augmented matrix**:

$$\left(\begin{array}{cccc|c} 1 & 2 & 1 & -2 & -1 \\ 2 & 3 & 3 & 3 & 10 \\ -1 & -4 & 1 & -4 & -15 \\ -2 & -1 & -5 & 3 & 5 \end{array}\right).$$

We apply **Gaussian elimination**, beginning with adding multiples of the first row to the second, third and fourth rows; specifically, the operations $R_2 \to -2R_1 + R_2$, $R_3 \to R_1 + R_3$ and $R_4 \to 2R_1 + R_4$:

$$\to \left(\begin{array}{cccc|c} 1 & 2 & 1 & -2 & -1 \\ 0 & -1 & 1 & 7 & 12 \\ 0 & -2 & 2 & -6 & -16 \\ 0 & 3 & -3 & -1 & 3 \end{array}\right).$$

Scale the second row by -1 to obtain:

$$\to \left(\begin{array}{cccc|c} 1 & 2 & 1 & -2 & -1 \\ 0 & 1 & -1 & -7 & -12 \\ 0 & -2 & 2 & -6 & -16 \\ 0 & 3 & -3 & -1 & 3 \end{array}\right).$$

We next add multiples of the second row to rows three and four. These operations are: $R_3 \to 2R_2 + R_3, R_4 \to (-3)R_2 + R_4$:

$$\to \left(\begin{array}{cccc|c} 1 & 2 & 1 & -2 & -1 \\ 0 & 1 & -1 & -7 & -12 \\ 0 & 0 & 0 & -20 & -40 \\ 0 & 0 & 0 & 20 & 39 \end{array}\right).$$

Scale the third row by $-\frac{1}{20}$:

$$\to \begin{pmatrix} 1 & 2 & 1 & -2 & | & -1 \\ 0 & 1 & -1 & -7 & | & -12 \\ 0 & 0 & 0 & 1 & | & 2 \\ 0 & 0 & 0 & 20 & | & 39 \end{pmatrix}.$$

Complete the **forward pass** by adding -20 times the third row to the fourth row:

$$\to \begin{pmatrix} 1 & 2 & 1 & -2 & | & -1 \\ 0 & 1 & -1 & -7 & | & -12 \\ 0 & 0 & 0 & 1 & | & 2 \\ 0 & 0 & 0 & 0 & | & -1 \end{pmatrix}.$$

The last column is a **pivot column** and therefore the system is **inconsistent**.

Exercises

In exercises 1-4, write down the matrix for each of the given systems. Indicate whether the matrix is augmented or not. See **Method** (1.2.1).

1.

$$\begin{array}{rcrcrcr} 2x & + & 2y & - & 3z & = & 0 \\ -x & + & 5y & - & z & = & 1 \end{array}$$

2.

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

3.

$$\begin{array}{rcrcrcr} \frac{1}{2}x & + & 2y & - & \frac{3}{4}z & = & 0 \\ x & - & \frac{1}{5}y & + & z & = & 0 \end{array}$$

4.

$$\begin{array}{rrrrcr} x_1 & +\,x_2 & -\,2x_3 & -\,2x_4 & = & 5 \\ -2x_1 & +\,2x_2 & -\,\frac{1}{2}x_3 & +\,3x_4 & = & -2 \\ x_1 & & +\,2x_3 & -\,x_4 & = & -1 \end{array}$$

In exercises 5-8 write a linear system which corresponds to the given matrix.
See **Method** (1.2.3).

5.

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 9 \end{pmatrix}$$

6.

$$\begin{pmatrix} 1 & -3 & 3 & | & 1 \\ 2 & -6 & 7 & | & 3 \end{pmatrix}$$

7.

$$\begin{pmatrix} 3 & -1 & 0 & 2 \\ -2 & -1 & 3 & -6 \\ 1 & 4 & -2 & 5 \\ 4 & 0 & 1 & 1 \end{pmatrix}$$

8.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & | & -1 \\ -1 & 0 & 1 & 2 & | & 2 \\ 1 & 4 & 1 & 4 & | & 3 \\ -1 & 0 & 1 & 8 & | & 0 \end{pmatrix}$$

In exercises 9-15 determine whether the given matrix is in **row echelon form**. If not, explain what condition fails.

9.

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 2 & 4 \end{pmatrix}$$

10.

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 0 \end{pmatrix}$$

11.

$$\begin{pmatrix} 2 & 4 & 8 \\ 0 & 3 & 6 \\ 0 & 0 & -1 \end{pmatrix}$$

12.

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

13.

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

14.

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

15.

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

16. Determine which, if any, of those matrices from exercises 9-15 is in **reduced echelon form**.

In exercises 17-20 apply **Gauss Elimination** to the given matrix in order to obtain its **reduced row echelon form**.

17.

$$\begin{pmatrix} 2 & -1 & 3 \\ 1 & -1 & 2 \\ 3 & -3 & 7 \end{pmatrix}$$

18.

$$\begin{pmatrix} 1 & -3 & 1 & 1 \\ 2 & -5 & 4 & -1 \\ -2 & 4 & 3 & -5 \\ 1 & 1 & 1 & -3 \end{pmatrix}$$

19.

$$\begin{pmatrix} 0 & 0 & 0 & 1 & -1 \\ 1 & 2 & -1 & 2 & 1 \\ 2 & 4 & -2 & 6 & 0 \\ 3 & 6 & -6 & 0 & 0 \end{pmatrix}$$

20.

$$\begin{pmatrix} 0 & -1 & 1 & 0 & 2 & 2 \\ 2 & -1 & -3 & -6 & -1 & -2 \\ -1 & 2 & 0 & -3 & 3 & 2 \\ 2 & 0 & -4 & 6 & 2 & -3 \\ 2 & 2 & -6 & 6 & 2 & 2 \end{pmatrix}$$

In exercises 21-24 the given matrices are in **reduced row echelon form** (check this). Assume each matrix corresponds to a **homogeneous linear system**. Write down the system and determine the general solution. See **Method** (1.2.2) and **Method** (1.2.4).

21.

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

22.

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

23.

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

24.

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

In exercises 25-27 the given matrices are in **reduced row echelon form** (check this). Assume each is the **augmented matrix** corresponding to a **inhomogeneous linear system**. Write down the system. State whether it is **consistent or inconsistent**. If it is consistent state how many **free variables** it contains and then determine the general solution. See **Method** (1.2.2) and **Method** (1.2.3).

25.

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

26.

$$\begin{pmatrix} 1 & -1 & 0 & 0 & -3 & | & 5 \\ 0 & 0 & 1 & 0 & 4 & | & -7 \\ 0 & 0 & 0 & 1 & -7 & | & 2 \end{pmatrix}$$

27.

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

In exercises 28-37 write down the **augmented matrix** of the given linear system. Find a **reduced echelon form** of this matrix. State whether the system is **consistent or inconsistent**. State the number of **free variables** and determine the general solution of the linear system. See **Method** (1.2.4).

28.

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

29.

$$\begin{array}{rcrcrcrcr} 4x_1 & - & x_2 & - & x_3 & - & x_4 & = & 0 \\ x_1 & - & x_2 & - & x_3 & + & 2x_4 & = & 0 \\ 3x_1 & - & x_2 & - & x_3 & - & x_4 & = & 0 \\ & & x_2 & - & 3x_3 & - & 2x_4 & = & 0 \end{array}$$

30.

$$\begin{array}{rcrcrcr} 2x & - & y & + & z & = & -3 \\ -3x & + & 3y & - & 2z & = & 11 \\ 4x & - & 3y & + & 6z & = & -14 \end{array}$$

31.

$$\begin{array}{rcrcrcr} x & + & 2y & + & 3z & = & 1 \\ 3x & + & 5y & + & 6z & = & -5 \\ 2x & + & 5y & + & 8z & = & -2 \end{array}$$

32.

$$\begin{array}{rcrcrcrcr} 2x_1 & - & x_2 & - & 3x_3 & - & 2x_4 & = & 1 \\ x_1 & - & x_2 & - & 4x_3 & - & 2x_4 & = & 5 \\ 3x_1 & - & x_2 & - & x_3 & - & 3x_4 & = & -2 \\ x_1 & & & + & 2x_3 & - & x_4 & = & -4 \end{array}$$

33.

$$\begin{array}{rcrcrcrcr} x_1 & - & x_2 & + & 3x_3 & - & x_4 & = & -1 \\ x_1 & + & 2x_2 & + & x_3 & + & x_4 & = & 5 \\ 2x_1 & + & 3x_2 & + & 3x_3 & + & x_4 & = & 9 \\ x_1 & + & x_2 & + & 2x_3 & & & = & 4 \end{array}$$

34.

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

35.

$$\begin{array}{rcrcrcrcr} x_1 & - & 2x_2 & - & 3x_3 & - & 4x_4 & = & -1 \\ -x_1 & + & 2x_2 & + & 4x_3 & - & 3x_4 & = & 11 \\ 2x_1 & - & 3x_2 & - & 2x_3 & - & 2x_4 & = & -5 \end{array}$$

36.

$$\begin{array}{rcrcrcrcr} 3x_1 & + & 5x_2 & - & x_3 & - & 2x_4 & = & -4 \\ x_1 & + & 2x_2 & + & 3x_3 & - & x_4 & = & 5 \\ 2x_1 & + & 3x_2 & - & 4x_3 & - & x_4 & = & -9 \end{array}$$

37.

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

In exercises 38-43 answer true or false and give an explanation.

38. If A and B are **row equivalent** and B and C are row equivalent then A and C are row equivalent.

39. If A is 3×4 matrix then a **linear system** with **augmented matrix** A either has a unique **solution** or is **inconsistent** .

40. Every matrix is **row equivalent** to a matrix in **echelon form** .

41. If A is an $m \times n$ matrix and every column is a **pivot column** then $m \leq n$.

42. If A and B are **row equivalent** and both are in **echelon form** then $A = B$.

43. If A and B are **row equivalent** and both are in **reduced echelon form** then $A = B$.

Challenge Exercises (Problems)

In 1-5 use **Theorem** (1.2.3) to draw conclusions about whether the **inhomogeneous linear system** with the given **augmented matrix** has no **solution**, a unique solution or infinitely many solutions.

1.

$$\begin{pmatrix} -2 & 3 & -4 & | & 0 \\ 0 & 4 & 2 & | & -1 \\ 0 & 0 & -2 & | & 3 \end{pmatrix}$$

2.

$$\begin{pmatrix} 4 & -3 & 2 & 0 & | & -1 \\ 0 & 2 & -1 & 4 & | & 0 \\ 0 & 0 & 3 & -2 & | & -5 \\ 0 & 0 & 0 & 4 & | & 2 \end{pmatrix}$$

3.

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

4.

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

5.

$$\left(\begin{array}{cccc|c} 3 & -2 & 1 & 0 & 5 \\ 0 & 0 & 4 & -1 & 2 \\ 0 & 0 & 0 & 7 & 14 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

6. Show that the following system of linear equations has (1,1,1,1) as a non-trivial solution and then explain why it has infinitely many solutions. See **Theorem** (1.2.4).

$$\begin{array}{ccccccccc} x_1 & - & x_2 & + & x_3 & - & x_4 & = & 0 \\ -3x_1 & + & x_2 & + & x_3 & + & x_4 & = & 0 \\ -x_1 & - & x_2 & + & x_3 & + & x_4 & = & 0 \\ x_1 & - & x_2 & - & x_3 & + & x_4 & = & 0 \end{array}$$

7. Explain why the following system must have infintely many solutions.

$$\begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & + & x_4 & = & 0 \\ x_1 & - & x_2 & + & x_3 & - & x_4 & = & 0 \\ 2x_1 & + & x_2 & + & x_3 & + & 2x_4 & = & 0 \end{array}$$

In each of 8-11 assume the given matrix is the **coefficient matrix** of a **homogeneous linear system**. Use **Theorem** (1.2.4) to determine whether the system has only the trivial solution or infinitely many solutions.

8.

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

9.

$$\begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & 5 & 4 & 3 \\ 0 & 0 & 6 & 5 \\ 0 & 0 & 0 & 7 \end{pmatrix}$$

10.

$$\begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

11.

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

Quiz Solutions

1. $(x, y, z) = (-1, -2, 3)$.

If you did not get this right you should see **Method** (1.1.1).

2.

$$\begin{array}{ccccccccc} x & + & y & + & z & + & 3w & = & 2 \\ & & y & & & - & 11w & = & 2 \\ & & & & z & + & 29w & = & 5 \end{array}$$

There are other correct forms but all should have the solution $(x, y, z, w) = (2, 2, 5, 0)$.

If you did not get this right you should see **Method** (1.1.2).

3. $(x, y, z, w) = (-6 - 2z + 5w, 4 - 2w, z, w)$.

If you did not get this right you should see **Method** (1.1.2).

1.3. How to Use it: Applications of Linear Systems

In this section we demonstrate several ways in which linear systems are used. These are:

Networks

Electric Circuits

Traffic Flow

Balancing Chemical Equations

Fitting Curves to Points

Blending Problems and Allocation of Resources

What You Need to Know

The following are concepts that you will need to know to understand the applications developed in this section.

linear equation

solution of linear equation

solution set of a linear equation

equivalent linear equations

standard form of a linear equation

leading term linear equation in standard form

leading variable of a linear equation in standard form

free variables of a linear equation in standard form

linear system

constants of a linear system

coefficients of a linear system

inhomogeneous linear system

homogenous linear system

solution of a linear system

solution set of a linear system

consistent linear system

inconsistent linear system

equivalent linear systems

echelon form for a linear system

leading variable

free variable

elementary equation operation

matrix

entries of a matrix

rows of a matrix

columns of a matrix

zero row of a matrix

non-zero row of a matrix

leading term of a non-zero row of a matrix

coefficient matrix of a linear system

augmented matrix of a linear system

elementary row operation of a matrix

row equivalent matrices

matrix in echelon form

pivot position of a matrix in echelon form

pivot column of a matrix in echelon form

matrix in reduced echelon form

echelon form of a matrix

reduced echelon form of a matrix

pivot positions of a matrix

pivot columns of a matrix

What You Need to Be Able to do

The following are the algorithms you need to have mastered in order to find solutions to the exercises of this section:

1. Given a **linear system**

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

write down the **corresponding matrix of the system**.

2. Given a matrix A write down the **homogeneous linear system** which corresponds to A. Given a matrix A write down the **inhomogeneous linear system** which corresponds to A.

3. Given a matrix A apply **Gaussian Elimination** to obtain an **echelon form** of A, the **reduced echelon form** of A.

4. Given a linear system determine if it is **consistent or inconsistent**. If a linear system it is consistent, find the general **solution** of the system.

Networks

Many situations can be modeled by the notion of a ***network***. The most general way to think of a network is as a collection of ***nodes*** (sometimes also called ***vertices*** or ***junctions***) with connections (often called ***arcs***, ***edges*** or ***branches***) between some of the pairs of the nodes. A connection indicates that there is some ***flow*** (goods, services, energy, traffic, information, electrical current) between the given pair of nodes. The connection is typically oriented, that is, given a direction, to indicate where the flow came from and where it is going.

A node is called a ***source*** if all the arcs connected to it are directed away from it and is a ***sink*** if all the arcs connected to it are directed towards it. See **Figure** (1.3.1). In Figure (1.3.1) node is a source, nodes 4 and 5 are sinks, while neither node 1 or node 3 is a source or a sink.

The Basic Assumption for Networks

In network analysis, the **basic assumption** is that every node which is neither a source nor a sink is ***balanced***: the sum of the flows entering a node is equal to the sum of the

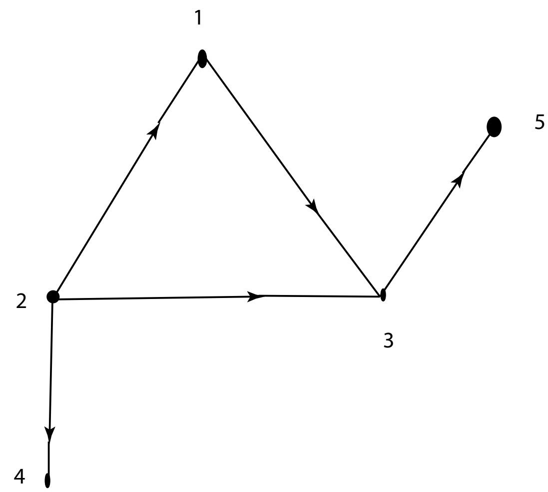


Figure 1.3.1: A typical network

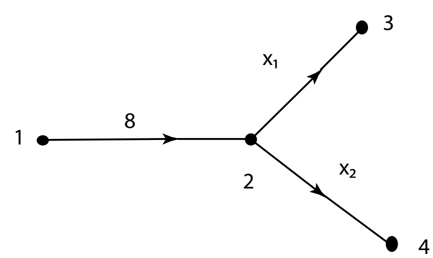


Figure 1.3.2: Balance in a network

flows leaving the node. We also say that the network is **in equilibrium**. See Figure (1.3.2).

In Figure (1.3.2) node 1 is a source and nodes 3 and 4 are sinks, while node 2 is neither a source nor a sink. At node 2 we get the equation $x_1 + x_2 = 8$ which follows from the basic assumption that we have equilibrium at node 2.

Given a network we can use the **basic assumption** to write down a **system of linear equations**, with one equation for each node which is neither a **source nor a sink** and a variable for each unknown flow. We illustrate with a couple of examples.

Example 1.3.1. Set up a **linear system** which represents the **network** shown in **Figure** (1.3.3). Assuming all the flows x_1, x_2, x_3, x_4 are non-negative find a solution to this network.

solution with x_4 as large as possible.

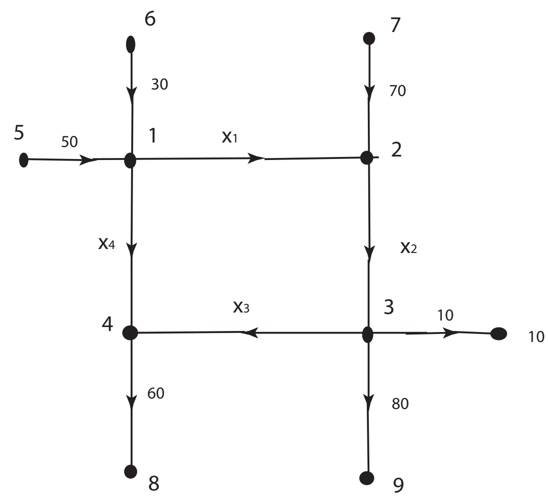


Figure 1.3.3: A ten node network

This network has three **sources**, namely, nodes 5, 6 and 7, and three **sinks**: nodes 8, 9 and 10. We require balance at nodes 1,2, 3 and 4. This is represented by the following system of equations:

$$\begin{aligned} x_1 + x_4 &= 80 \text{ at node } 1 \\ x_1 + 70 &= x_2 \text{ at node } 2 \\ x_3 + 90 &= x_2 \text{ at node } 3 \\ x_3 + x_4 &= 60 \text{ at node } 4 \end{aligned}$$

The **matrix of this system** is

$$\begin{pmatrix} 1 & 0 & 0 & 1 & | & 80 \\ 1 & -1 & 0 & 0 & | & -70 \\ 0 & 1 & -1 & 0 & | & 90 \\ 0 & 0 & 1 & 1 & | & 60 \end{pmatrix}$$

Using **Gauss elimination** we obtain the **reduced echelon form**.

$$\begin{pmatrix} 1 & 0 & 0 & 1 & | & 80 \\ 0 & 1 & 0 & 1 & | & 150 \\ 0 & 0 & 1 & 1 & | & 60 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$$

The **linear system** which has this (augmented matrix) as it **matrix** has as its general solution the vector

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 80 - t \\ 150 - t \\ 60 - t \\ t \end{pmatrix}.$$

Since it is required that each of the x_i be non-negative we must have $t \leq 60$. Since $x_4 = t$ the solution with x_4 as large as possible is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 20 \\ 90 \\ 0 \\ 60 \end{pmatrix}.$$

Example 1.3.2. Set up a **linear system** to represent the **network** shown in **Figure** (1.3.4). Find the **general solution to the system**.

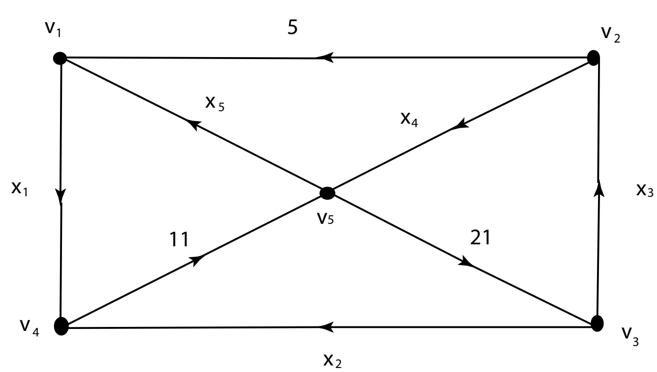


Figure 1.3.4: A five node network

There are no **sinks or sources** in this network and so we get five equations in five variables:

$$\begin{aligned} x_5 + 5 &= x_1 \text{ at node } 1 \\ x_3 &= x_4 + 5 \text{ at node } 2 \\ 21 &= x_2 + x_3 \text{ at node } 3 \\ x_1 + x_2 &= 11 \text{ at node } 4 \\ x_4 + 11 &= x_5 + 21 \text{ at node } 5 \end{aligned}$$

the matrix for this system is the following:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & -1 & | & 5 \\ 0 & 0 & 1 & -1 & 0 & | & 5 \\ 0 & 1 & 1 & 0 & 0 & | & 21 \\ 1 & 1 & 0 & 0 & 0 & | & 11 \\ 0 & 0 & 0 & 1 & -1 & | & 10 \end{pmatrix}$$

Using **Gaussian elimination** we obtain the following **reduced echelon form** of this matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & -1 & | & 5 \\ 0 & 1 & 0 & 0 & 1 & | & 6 \\ 0 & 0 & 1 & 0 & -1 & | & 15 \\ 0 & 0 & 0 & 1 & -1 & | & 10 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$$

The **linear system** which has this augmented matrix is as follows:

$$\begin{array}{ccccccc} x_1 & & & & -\, x_5 & = & 5 \\ & x_2 & & & +\, x_5 & = & 6 \\ & & x_3 & & -\, x_5 & = & 15 \\ & & & x_4 & -\, x_5 & = & 10 \end{array}$$

Setting $x_5 = t$ we obtain the general solution $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 5+t \\ 6-t \\ 15+t \\ 10+t \\ t \end{pmatrix}$.

The general notion of a **network** can be used to describe electric circuits, automobile traffic on a city's streets, water through irrigation pipes, the exchange of goods and services in an economy, the interchange of nutrients and minerals in an ecological niche and many other possibilities. We illustrate this with some examples.

Electric Circuits

An electric circuit consists of three elements: ***voltage sources*** (batteries), ***resistors*** (lights, appliances) and ***conductors*** (wires). A voltage source provides an ***electro-motive force*** E which is measured in **volts** (V) and **pushes electrons through the network**. The electrons **flow from the positive to the negative and this creates a potential difference**. ***Current***, denoted by I, is the **rate at which the electrons flow through a network** and is measured in ***amperes*** (A). Finally, **resistors use up the current (lost as heat), and lower the voltage**. ***Resistance*** is represented by R and measured in **ohms** (Ω). An electric network is governed by **Kirchoff's laws**:

Kirchoff's Voltage Law (Conservation of Energy): The voltages around a closed path in a circuit must sum to zero. The voltages going from the negative to the positive terminal of a battery are positive while those arising from a current passing through a resistor are negative.

Kirchoff's Current Law (Conservation of Current) The sum of currents entering a node is equal to the sum of currents leaving a node.

We also needs **Ohm's law** which tells us what the voltage drop is when a current passes through a resistor:

Ohm's Law

The drop in potential difference E across a resistor is given by IR where I is the current measured in amps and R is the resistance measured in ohms.

Applying these to an electric circuit will give rise to a system of linear equations. We do an example.

Example 1.3.3.

Consider the **electric circuit** of **Figure** (1.3.5). Find the currents, I_1, I_2, I_3 in each branch of the network.

Using **Kirchoff's** and **Ohm's** laws we obtain the following equations:

$I_1 - I_2 + I_3 = 0$, from nodes B and D

$4I_1 - 6I_3 = 18$, going around loop $ADBD$

$8I_2 + 6I_3 = 16$, going around loop $CBDB$

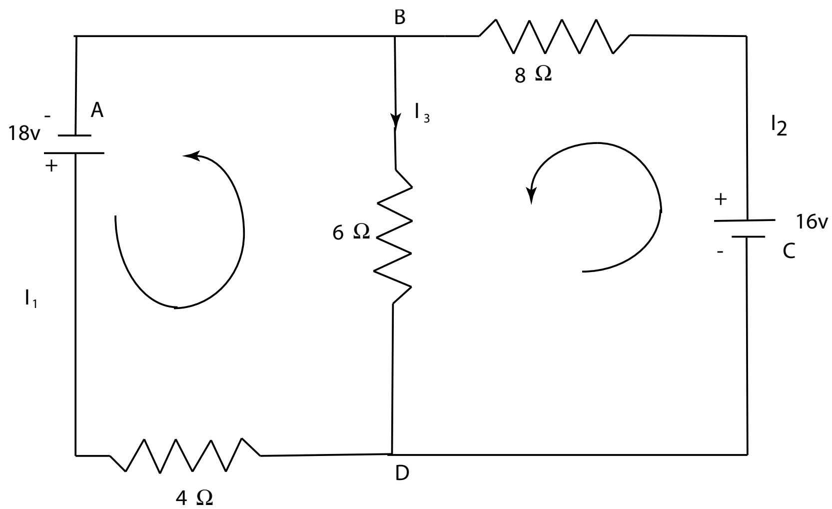


Figure 1.3.5: An example of an electric circuit

$4I_1 + 8I_2 = 34$, going around loop $ADCBA$

This **linear system** has the following **matrix**:

$$\begin{pmatrix} 1 & -1 & 1 & | & 0 \\ 4 & 0 & -6 & | & 18 \\ 0 & 8 & 6 & | & 16 \\ 4 & 8 & 0 & | & 34 \end{pmatrix}$$

Using **Gaussian elimination** we obtain the following **reduced echelon form** for this matrix:

$$\begin{pmatrix} 1 & 0 & 0 & | & \frac{87}{26} \\ 0 & 1 & 0 & | & \frac{67}{26} \\ 0 & 0 & 1 & | & -\frac{10}{13} \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

Thus, there is a unique **solution**. Note that the negative sign for I_3 means that the current flows in the direction opposite that indicated in **Figure** (1.3.5).

Traffic flow

Networks and **linear systems** can be used to model ***traffic flows*** as the next simplified example illustrates:

Example 1.3.4. **Figure** (1.3.6) represents four adjacent intersections of one way streets in the center of a small town. Assume that the average number of cars entering and leaving these intersections over one hour intervals at midday are as labeled as in the figure. Find the amount of traffic between each of the intersections.

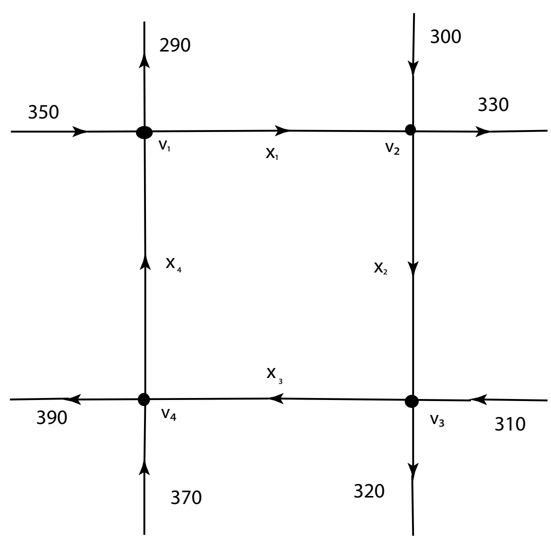


Figure 1.3.6: A traffic network

Assuming that the number of cars entering each intersection is equal to the number of cars exiting (no car stops and parks on the streets), we obtain the following equations:

$x_4 + 350 = x_1 + 290$, equivalently, $x_1 - x_4 = 60$ at intersection v_1

$x_1 + 300 = x_2 + 330$, equivalently, $x_1 - x_2 = 30$ at intersection v_2

$x_2 + 310 = x_3 + 320$, equivalently, $x_2 - x_3 = 10$ at intersection v_3

$x_3 + 370 = x_4 + 390$, equivalently, $x_3 - x_4 = 20$ at intersection v_4.

The **matrix of this system** is

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 60 \\ 1 & -1 & 0 & 0 & 30 \\ 0 & 1 & -1 & 0 & 10 \\ 0 & 0 & 1 & -1 & 20 \end{array}\right)$$

Using **Gaussian elimination** we obtain the following **reduced echelon form**:

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 60 \\ 0 & 1 & 0 & -1 & 30 \\ 0 & 0 & 1 & -1 & 20 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

As can be seen from the matrix, a priori, there are infinitely many **solutions** to this **linear system**. Setting $x_4 = t$ these are $x_1 = 60+t, x_2 = 30+t, x_3 = 20+t, x_4 = t$. However, each x_1, x_2, x_3, x_4 must be a non-negative integer which also implies that t is a non-negative integer.

Balancing Chemical Equations

Chemical reactions can be described by equations. Such an equation shows the chemicals that react, the ***reactants***, on the left hand side of the equation and the chemicals that result, the ***products***, on the right hand side. The chemicals are usually represented by their symbols. Instead of an equal sign between the two sides of this equation they are related by an arrow indicating that the reactants combine to make the products. To be ***balanced*** the equations must conserve mass which means that the number of molecules of each element on either side of the equation must be equal. This leads to a **linear system of equations**.

Example 1.3.5. Leavening is an important process in the preparation of baked products. One way of achieving this is through a chemical reaction between an acidic and an alkaline ingredient to produce carbon dioxide. This is often achieved through the use of a baking powder which is a complete mixture of at least one acid salt and an alkalilne. One such baking powder consists of monocalcium phosphate ($CaH_4(PO_4)_2$) and bicarbornate of soda ($NaHCO_3$). These react to produce tricalcium phosphate ($Ca_3(PO_4)_2$), disodium phosphate (Na_2HPO_4), carbon dioxide (CO_2) and water (H_2O). This is represented by the chemical equation

$$CaH_4(PO_4)_2 + NaHCO_3 \to Ca_3(PO_4)_2 + Na_2HPO_4 + CO_2 + H_2O$$

However, this qualitative equation does not tell us how much of each reactant is needed and the relative amounts of the product that are obtained. This is determined by application of linear algebra. In chemistry a standard unit of measurement is a **mole**: the amount of a substance which contains as many elementary entities (atoms, molecules, ions, electrons or other particles) as there are in 0.012 kilograms of carbon 12. We assume that we are mixing x_1 moles of monocalcium phosphate and x_2 moles of bicarbonate of soda and that produces x_3 moles tricalcium phosphate, x_4 moles of disodium phosphate, x_5 moles of carbon dioxide and x_6 moles of water. This is represented by the equation

$$x_1 CaH_4(PO_4)_2 + x_2 NaHCO_3 \rightarrow$$

$$x_3 Ca_3(PO_4)_2 + x_4 Na_2HPO_4 + x_5 CO_2 + x_6 H_2O \tag{1.14}$$

The number of each atom in (1.14) must balance before and after the reaction. This gives rise to the following equations:

$x_1 = 3x_3$ from balancing the number of Ca (calcium) atoms

$4x_1 + x_2 = x_4 + 2x_6$ from balancing the number of H (hydrogen) atoms

$2x_1 = 2x_3 + x_4$ from balanciing the number of P (phosphate) atoms

$8x_1 + 3x_2 = 8x_3 + 4x_4 + 2x_5 + x_6$ from balancing the number of O (oxygen) atoms

$x_2 = 2x_4$ from balancing the number of Na (sodium) atoms

$x_2 = x_5$ from balancing the number of C (carbon) atoms

This is a **homogenous linear system** and has the following **matrix**:

$$\begin{pmatrix} 1 & 0 & -3 & 0 & 0 & 0 \\ 4 & 1 & 0 & -1 & 0 & -2 \\ 2 & 0 & -2 & -1 & 0 & 0 \\ 8 & 3 & -8 & -4 & -2 & -1 \\ 0 & 1 & 0 & -2 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \end{pmatrix}$$

Applying **Gaussian elimination** we obtain the following **reduced echelon form**:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & -\frac{3}{8} \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & -\frac{1}{8} \\ 0 & 0 & 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Setting $x_6 = t$ we get the **general solution**. However, because of the physical nature of this problem our solutions must all be non-negative integers. The smallest value of t for which this is true is $t = 8$ which gives the solution $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \\ 1 \\ 4 \\ 8 \\ 8 \end{pmatrix}$.

All other solutions are multiples of this one. Thus, the "smallest" equation which balances this reaction is

$$3CaH_4(PO_4)_2 + x8NaHCO_3 \to Ca_3(PO_4)_2 + 4Na_2HPO_4 + 8CO_2 + 8H_2O$$

Fitting a Curve to Points

It is often the case that collected data is in the form of a collection of points (x_1, y_1), $(x_2, y_2), \dots, (x_n, y_n)$ and one wants to find a polynomial of least degree whose graph contains these points. This is called ***polynomial curve fitting***. If no two of the points lie on a vertical line, equivalently, for $i \neq j, x_i \neq x_j$ then it can be shown that there is a unique polynomial $f(x)$ of degree at most $n-1$, $f(x) = a_0 + a_1 x + \dots a_{n-2}x^{n-2} + a_{n-1}x^{n-1}$ whose graph contains each of the points $(x_i, y_i), i = 1, 2, \dots, n-1$.

To find the coefficients of this polynomial substitute each point in the expression for $f(x)$. This will give a **linear equation** in $a_0, a_1, \dots, a_{n-1}$. The n points (x_i, y_i) then gives a **linear system** of n equations in these n variables. The system looks like the following:

$$a_0 + a_1 x_1 + \dots + a_{n-2}x_1^{n-2} + a_{n-1}x_1^{n-1} = y_1$$

$$a_0 + a_1 x_2 + \dots + a_{n-2}x_2^{n-2} + a_{n-1}x_2^{n-1} = y_2$$

$$\vdots$$

$$a_0 + a_1 x_n + \dots + a_{n-2}x_n^{n-2} + a_{n-1}x_n^{n-1} = y_n$$

The **matrix** of this linear system is

$$\left(\begin{array}{ccccc|c} 1 & x_1 & \dots & x_1^{n-2} & x_1^{n-1} & y_1 \\ 1 & x_2 & \dots & x_2^{n-2} & x_2^{n-1} & y_2 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 1 & x_n & \dots & x_n^{n-2} & x_n^{n-1} & y_n \end{array}\right)$$

It will be shown in the chapter on determinants that this system always has a unique **solution**. We illustrate with a couple of examples.

Example 1.3.6. Find the polynomial $f(x) = a_2x^2 + a_1x + a_0$ of degree at most two which contains the points (1,1), (2, 3) and (3, 7).

The **linear system** we obtain by substituting the points into $f(x)$ is as follows:

$$a_0 + a_1 + a_2 = 1$$

$$a_0 + 2a_1 + 4a_2 = 3$$

$$a_0 + 3a_1 + 9a_2 = 7$$

The **augmented matrix** of this **linear system** is

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 \\ 1 & 3 & 9 & 7 \end{array}\right)$$

The **reduced echelon form** of this matrix is

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

Therefore the polynomial is $x^2 - x + 1$. The graph of this quadratic is shown in **Figure** (1.3.7).

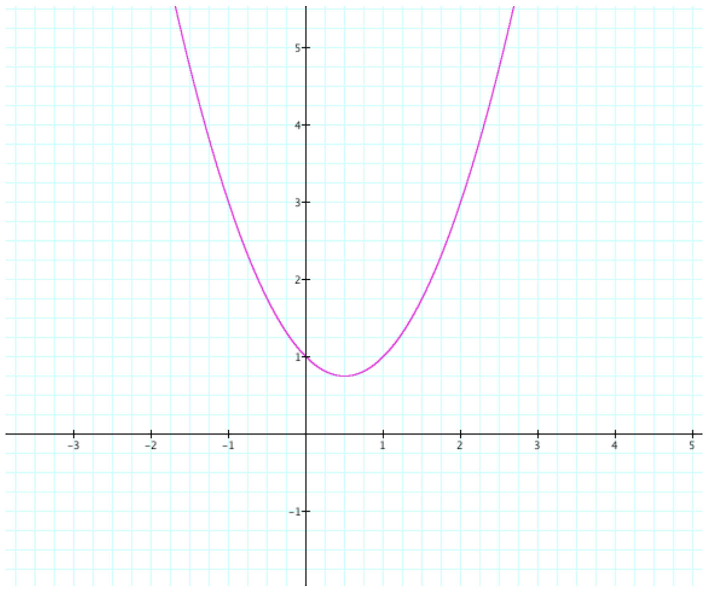


Figure 1.3.7: Graph of $y = x^2 - x + 1$

Example 1.3.7. Find the unique polynomial $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ of degree at most three whose graph contains the points (-1,2), (1,4), (2,5) and (3, -2)

Substituting the points into $f(x)$ we obtain the following **linear system**:

$$\begin{array}{ccccccccc} a_0 & - & a_1 & + & a_2 & - & a_3 & = & 2 \\ a_0 & + & a_1 & + & a_2 & + & a_3 & = & 4 \\ a_0 & + & 2a_1 & + & a_2 & + & a_3 & = & 5 \\ a_0 & + & 3a_1 & + & 9a_2 & + & 27a_3 & = & -2 \end{array}$$

The **augmented matrix** of this **linear system** is

$$\begin{pmatrix} 1 & -1 & 1 & -1 & | & 2 \\ 1 & 1 & 1 & 1 & | & 4 \\ 1 & 2 & 4 & 8 & | & 6 \\ 1 & 3 & 9 & 27 & | & -2 \end{pmatrix}$$

The **reduced echelon form** of this matrix, obtained using **Gaussian elimination** , is

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

The required polynomial is $f(x) = 1 + 2x + 2x^2 - x^3$. The graph of this polynomial is shown in **Figure** (1.3.8).

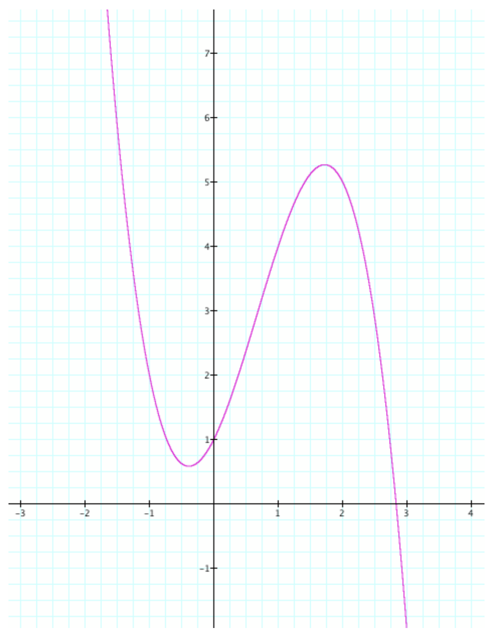


Figure 1.3.8: Graph of $y = -x^3 + 2x^2 + 2x + 1$

Blending Problems and Allocation of Resources

Many problems arise where a desired product must meet several specific numerical criteria and there are multiple sources of ingredients that can be mixed together. For example, a simple diet for a person (or an animal) must contain a certain number of grams of protein, carbohydrates and fats while these are present in different proportions in the many foods that may be part of the diet. Similar problems arise in manufacturing. The following is an example:

Example 1.3.8. A firm wishes to market bags of lawn fertilizer which contain 25% nitrogen, 7% phosphoric acid and 8% potash. The firm has four chemical precursors C_1, C_2, C_3, C_4 which are to be combined to make the fertilizer. The percentage of each ingredient in a pound of these chemicals is given in the following table:

	C_1	C_2	C_3	C_4
Nitrogen	0.20	0.25	0	0.30
Phosphoric Acid	0.12	0.05	0.06	0.07
Potash	0	0.05	0.15	0.10

How much of each chemical should be mixed to obtain 100 pounds of fertilizer meeting these criteria?

Let x_i be the number of pounds of chemical C_i is used. Then since the total is to be 100 pounds we have the equation

$$x_1 + x_2 + x_3 + x_4 = 100.$$

Now x_1 pounds of chemical C_1 contains $0.20x_1$ pounds of nitrogen, x_2 pounds of C_2 contains $0.25x_2$ pounds of nitrogen, x_3 pounds of C_3 contains $0x_3 = 0$ pounds of nitrogren and x_4 pounds of C_4 contains $0.30x_4$ pounds of nitrogen. Since there are to be $0.25 \times 100 = 25$ pounds of nitrogen in the mixture we obtain the following **linear equation**.

By this reasoning the total number of pounds of nitrogen will be

$$0.20x_1 + 0.25x_2 + 0x_3 + 0.30x_4 = 25 \tag{1.15}$$

In a similar way we get an equation for phosphoric acid

$$0.12x_1 + 0.05x_2 + 0.06x_3 + 0.07x_4 = 7 \tag{1.16}$$

and we get the following equation for potash:

$$0x_1 + 0.05x_2 + 0.15x_3 + 0.10x_4 = 8 \tag{1.17}$$

Thus, we have a **linear system** with the following **augmented matrix**:

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 100 \\ 0.20 & 0.25 & 0 & 0.30 & 25 \\ 0.12 & 0.05 & 0.06 & 0.07 & 7 \\ 0 & 0.05 & 0.15 & 0.10 & 8 \end{array}\right)$$

Using **Gaussian elimination** we obtain the following **reduced echelon form** of this matrix. T

$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & 17.75 \\ 0 & 1 & 0 & 0 & | & 25.08 \\ 0 & 0 & 1 & 0 & | & 8.57 \\ 0 & 0 & 0 & 1 & | & 54.60 \end{pmatrix}$$

which gives us our **solution**, $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 17.75 \\ 25.08 \\ 8.57 \\ 54.60 \end{pmatrix}$.

Exercises

In exercises 1 - 6 find a **polynomial** $f(x)$ whose graph contains the given points. Sketch the graph and show the points on the graph.

1. (1,0), (2,0), (3,4)

2. (1,1), (2,5), (3,13)

3. (1,-1), (2,-1), (3,3)

4. (-1,-4), (0,-1), (1,0), (2,5)

5. (0,5), (1,2), (2,1), (3,2)

6. (-1,0), (0,0), (1,0), (2,3)

The exercises continue on the next page.

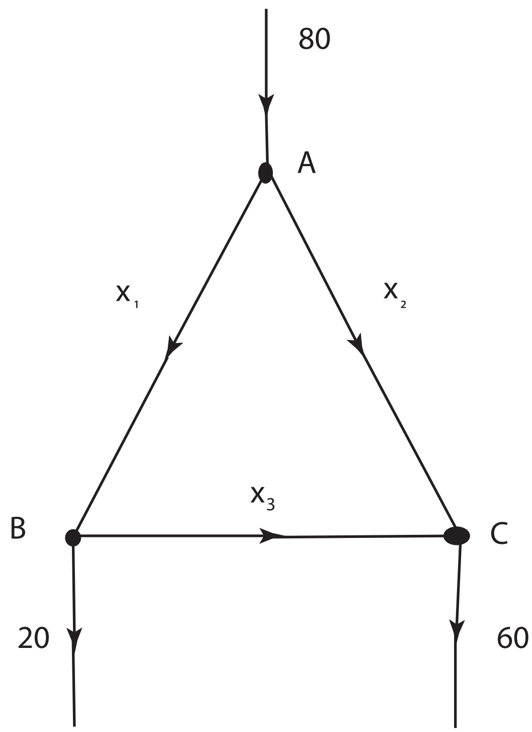


Figure 1.3.9: Exercise 7

7. The network shown in **Figure** (1.3.9) shows the flow of pedestrian traffic on several streets over a five minute interval. If the system is in **equilibrium** find the values of x_1, x_2 and x_3 in number of people per minute.

8. **Figure** (1.3.10) shows the flow of automobile traffic in the center of town over one hour. Assuming the system is in equilibrium determine the values of x_1, x_2, x_3, x_4 in number of cars per hour. Find the minimum and maximum flow in branch BC. Assume that all the streets are one way.

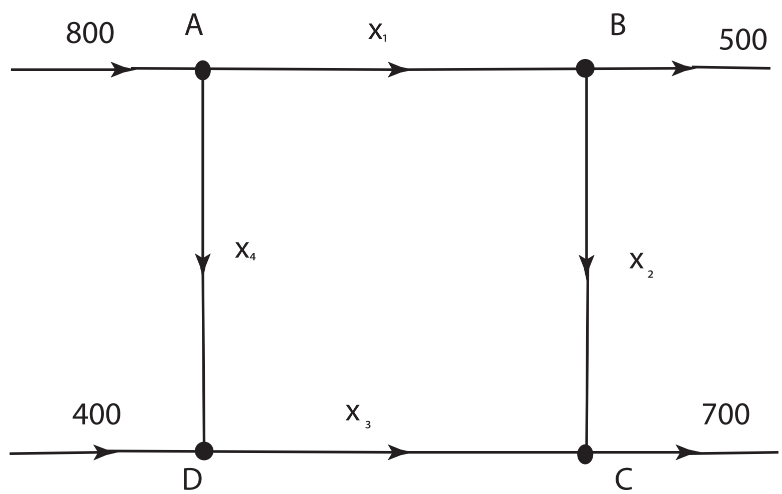


Figure 1.3.10: Exercise 8

9. **Figure** (1.3.11) represents the flow of water through a series of pipes. All the pipes are designed so that water flows in only one direction. The flows (in thousand gallons per minute) are as shown. Find the rate of flow in each branch of the network. If there is no flow in branch BC, then find the maximum flow in branch BD.

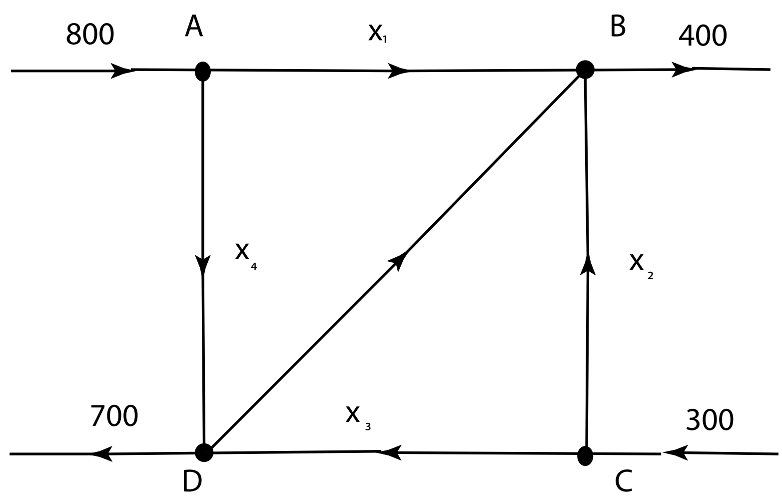


Figure 1.3.11: Exercise 9

10. **Figure** (1.3.12) shows the flow of bicycle traffic on a collection of one way paths. The number of bikes per hour is shown. Determine the flows on each path and determine the minimum flow in path CD.

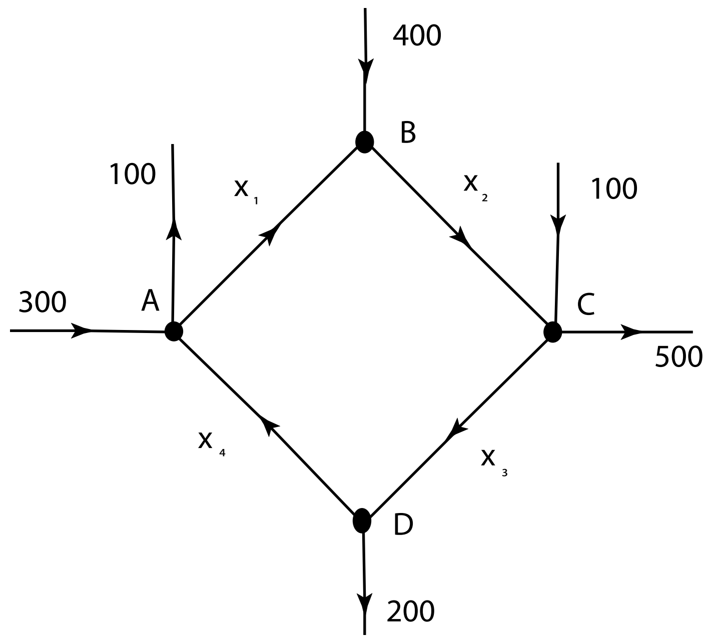


Figure 1.3.12: Exercise 10

11. **Figure** (1.3.13) shows a network flowing in **equilibrium**. The flows are as given and are in units per minute. Find the flows in each branch of the network and determine the minimum flow in branch AB and maximum flow in branch EF.

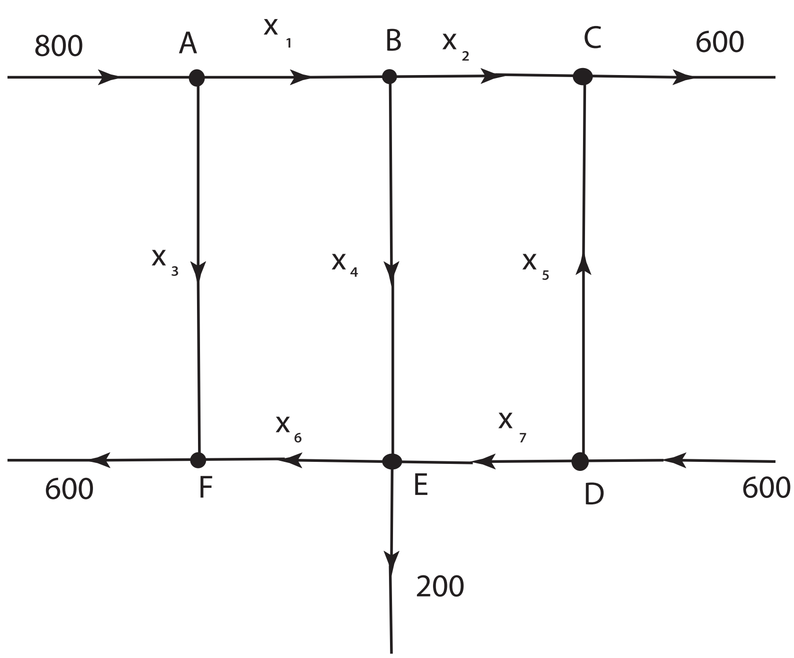


Figure 1.3.13: Exercise 11

12. **Figure** (1.3.14) shows a network in **equilibrium**. The flows are given in units per hour. Find the flows in each branch. Also determine the maximum and minimum flow in each branch of the network.

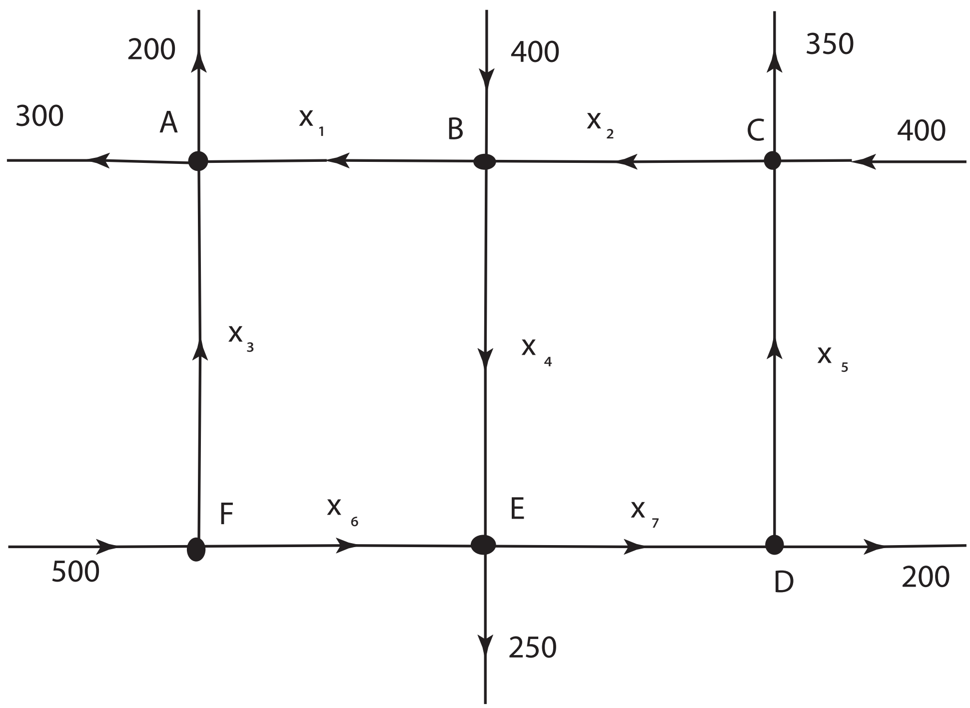
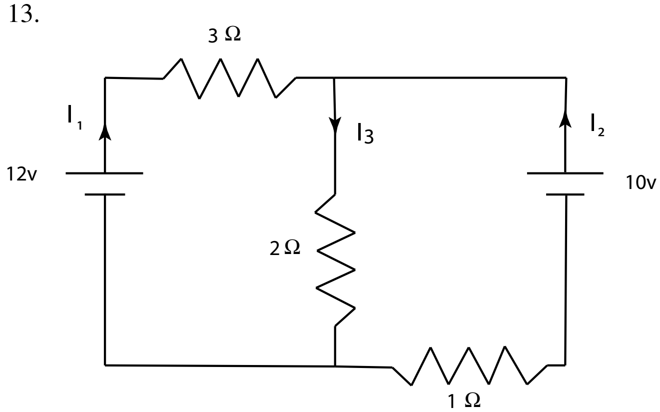


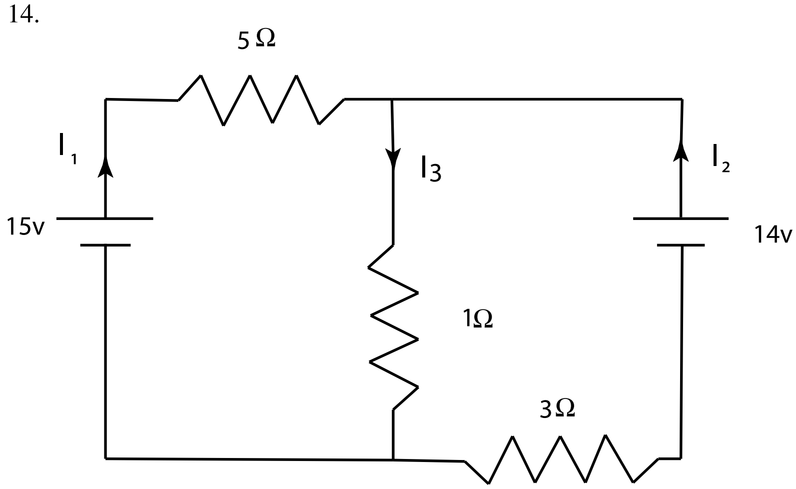
Figure 1.3.14: Exercise 12

In each of exercises 13 - 18 determine the value of the unknown currents in each **electric circuit**.

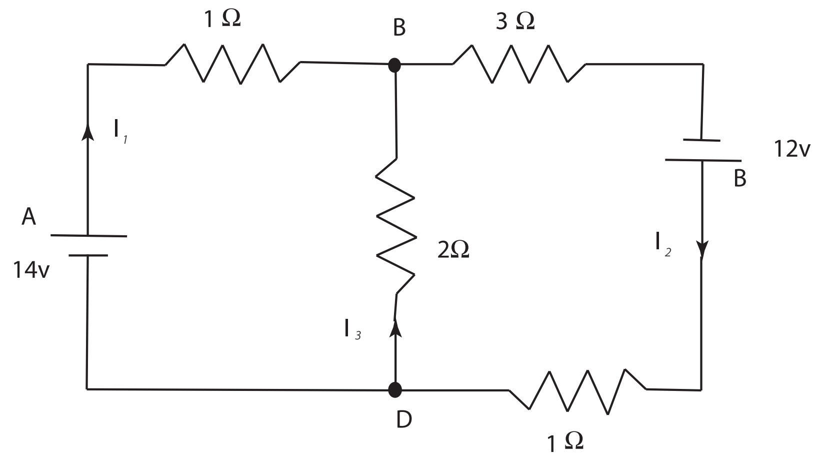
13.



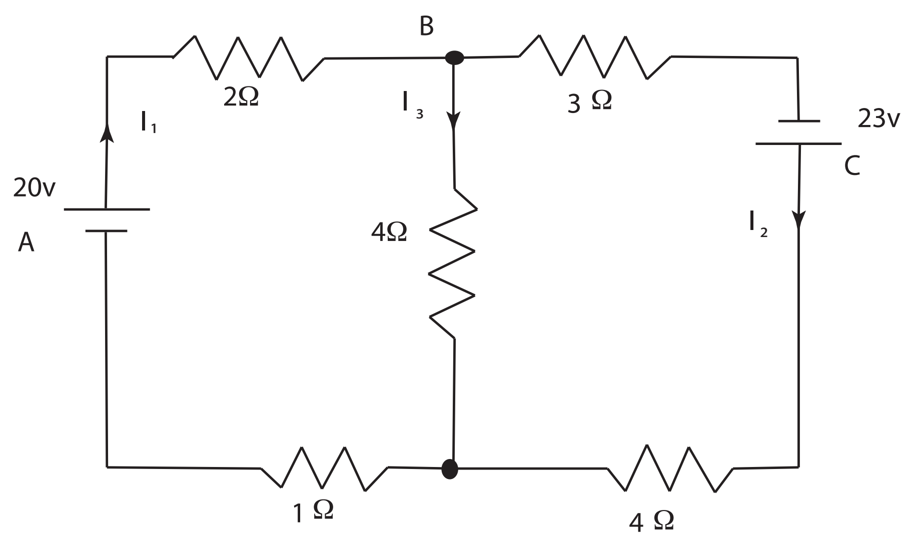
14.



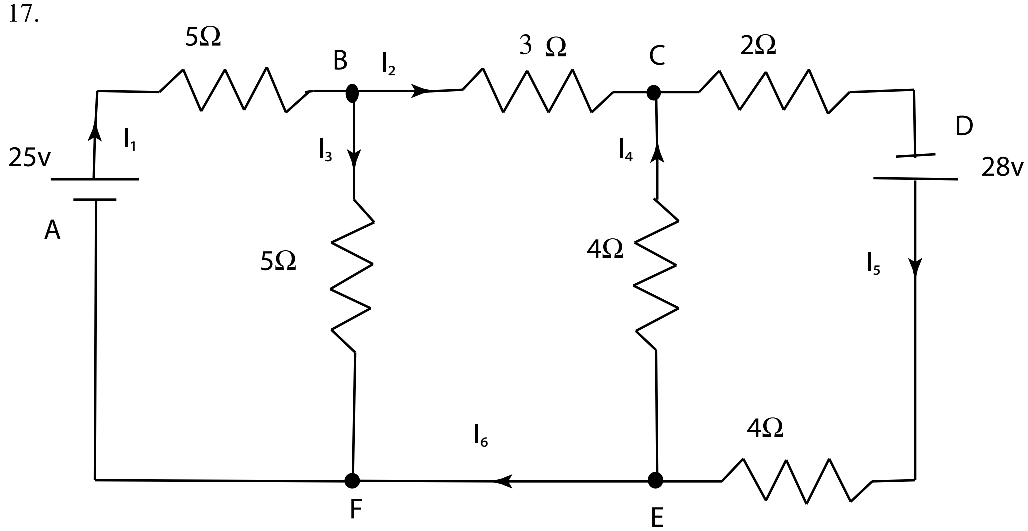
15.



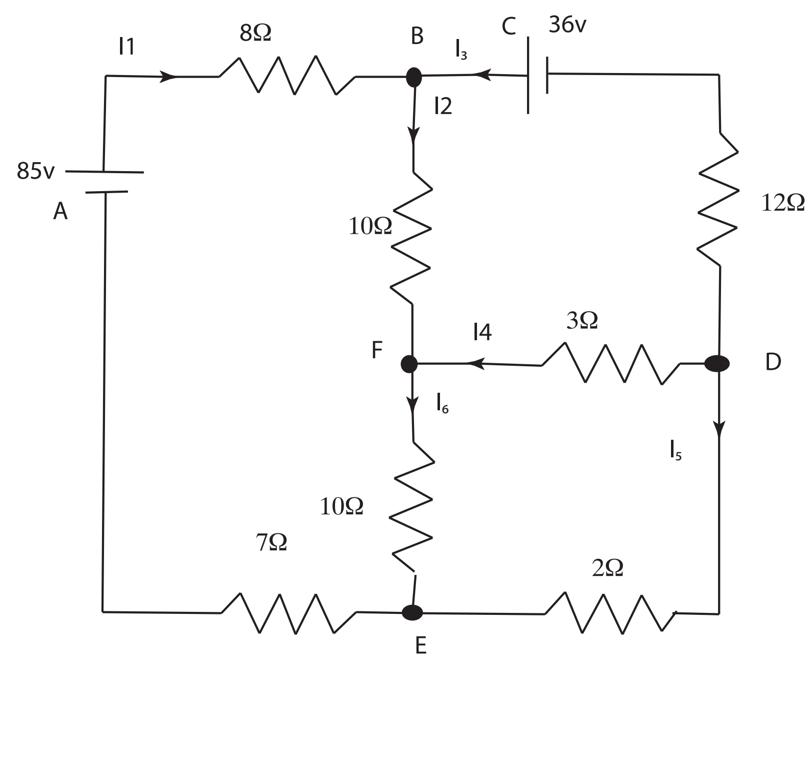
16.



17.


5Ω
B
I_2
3 Ω
C
2Ω
D
28v
25v
I_1
I_3
I_4
A
5Ω
4Ω
I_5
I_6
4Ω
F
E

18.


I1
8Ω
B
I_3
C
36v
I2
85v
A
10Ω
12Ω
I4
3Ω
F
D
I_6
I_5
10Ω
7Ω
2Ω
E

In exercises 19-23 balance the given **chemical equation**.

19. Methyl bromide combines with potassium hydroxide (potash) to produce methyl alcohol and potassium bromide:

$$CH_3Br + KOH \to CH_3OH + KBr$$

20. When oxygen is in abundance it will combine with organic compounds to produce carbon dioxide and water as it products. An example is ethanol combined with oxygen as represented by the formula

$$C_2H_6O + O_2 \to CO_2 + H_2O.$$

21. $KHC_8H_4O_4 + KOH \to K_2C_8H_4O_4 + H_2O$

22. $K_4Fe(CN)_6 + H_2SO_4 + H_2O \to K_2SO_4 + FeSO_4 + (NH_4)_2SO_4 + CO$

23. $C_6H_5Cl + SiCl_4 + Na \to (C_6H_5)_4Si + NaCl$

24. This exercise continues **Example** (1.3.8). Assume that a fifth chemical precursor is available for manufacturing fertilizer, C_5, which is 24% nitrogen, 4% phosphoric acid and 10% potash. Determine all possible solutions for combining $C_1 - C_5$ to obtain the 10,000 pounds of fertilizer. Also, determine which, if any, chemicals apart from C_5 are inessential, that is, are not needed to make a mixture.

25. A refinery has two sources of raw gasoline, $RT1$ and $RT2$ and wishes to use these to produce regular gasoline (RG) and premium gasoline (PG). $RT1$ has a performance rating of 110 and $RT2$ has a performance rating of 80. There are 40,000 gallons of $RT1$ and 20,000 gallons of $RT2$. Premium gas should have a performance rating of 100 and regular a performance rating of 90. Find how much of each of the raw gasoline types should be allocated to the production of premium and regular gas.

Chapter 2

The Vector Space $\mathbb{R}^n$

2.1. Introduction to Vectors: Linear Geometry

In this chapter we introduce and study one of the fundamental objects of linear algebra - vectors in $\mathbb{R}^n$. We begin in this section with an intuitive geometric approach to vectors in 2-space (the plane) and 3-space which will motivate the algebraic definition of vectors and the space $\mathbb{R}^n$ introduced in the next section. It is intended to give the reader an overview of some of the questions investigated in linear algebra as well as a geometric picture that may help you interpret and better understand the algebra. On the other hand, in subsequent sections we do not actually make use of the material in this section and it can be passed over. Therefore, if you like, you can proceed directly to the next section.

Am I Ready for This Material

Readiness Quiz

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

The following are concepts and results from Euclidean geometry and analytic geometry that you need to be familiar with in this section:

Cartesian coordinate system

line

ray

plane

coordinates of a point

line segment

distance formula

Pythagorean theorem

length of a line segment

parallel lines

parallelogram

congruent triangles

similar triangles

measure of an angle

Quiz

1. Find the length of the line segment $\overline{PQ}$ where P has coordinates (-2, 4) and Q has coordinates (2,7).

2. Find the length of the line segment $\overline{PQ}$ where P has coordinates (-3, 0, 2) and Q has coordinates (5, 4, 1).

3. If $P = (1, -2), Q = (3, 1)$ and $R = (0, -2)$ find S such that the line segments $\overline{PQ}$ and $\overline{RS}$ has the same length and are parallel.

Solutions

New Concepts

The material introduced in this section is essentially for motivational purposes. Though it will not be emphasized it is a good idea to become familiar with these concepts as they will facilitate a better understanding of the algebra developed in the subsequent sections.

directed line segment

length of a directed line segment

equivalence of directed line segments

geometric vector

canonical representative of a geometric vector

geometric definition of zero vector

geometric definition of the negative of a vector

geometric sum of vectors

scalar multiple of a geometric vector

span of two geometric vectors

Theory (Why it Works)

We begin with a geometric definition of vector from which we will extract a corresponding algebraic interpretation. We will do this in 2-space (the plane) and 3-space. Before giving our geometric definition of a vector, we review some concepts from the geometry of 2-space and 3-space (which we assume has a **Cartesian coordinate system**, that is, there are coordinate axes which are perpendicular to each other). We will denote by $\mathbf{O}$ the origin, which is either (0,0) when we are in the 2-space, or (0,0,0) in 3-space.

We assume familiarity with the concepts of **line** and **ray**. Recall that in 2-space two distinct lines $\mathcal{L}$ and $\mathcal{M}$ are said to be ***parallel*** if they do not meet. In 3-space the lines are parallel if they lie in a common plane and do not meet.

When $\mathcal{L}$ and $\mathcal{M}$ are parallel we will write $\mathcal{L} \parallel \mathcal{M}$. By convention, we will also consider any line $\mathcal{L}$ to be parallel to itself. Parallelism of lines satisfies the following properties:

(Reflexive Property) Any line $\mathcal{L}$ is parallel to itself, $\mathcal{L} \parallel \mathcal{L}$.

(Symmetric Property) If $\mathcal{L}$ is parallel to $\mathcal{M}$ then $\mathcal{M}$ is parallel to $\mathcal{L}$.

(Transitive Property) If $\mathcal{L}$ is parallel to $\mathcal{M}$ and $\mathcal{M}$ is parallel to $\mathcal{N}$ then $\mathcal{L}$ is parallel to $\mathcal{N}$.

We refer to all the lines parallel to a given line $\mathcal{L}$ as the ***parallel class of*** $\mathcal{L}$.

Every line has two **orientations** - starting at any point on a representative $\mathcal{L}$ of the class there are two ways that an imaginary being can go. By a ***direction*** we will mean a **parallel class** along with an orientation. This is illlustrated in Figure (2.1.1).

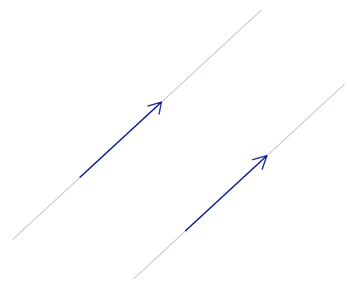


Figure 2.1.1: Direction of a line

We also recall that for two points P and Q in space we can define the ***distance***, $dist(P,Q)$, between them. This is computed from the coordinates of the points using the **distance formula** which is derived from the **Pythagorean theorem**. In 2-space this is given by

$$dist((p_1,p_2),(q_1,q_2)) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2}$$

and in 3-space by

$$dist((p_1,p_2,p_3),(q_1,q_2,q_3)) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+(q_3-p_3)^2}$$

We have now laid the foundation for the definition of our geometric vectors. The first element of this is a directed line segment from a point P to a point Q.

Definition 2.1. Given a pair of points (P,Q) the ***directed line segment***, $\vec{PQ}$, is the straight line segment joining P to Q with an arrow pointing from P to Q. P is the ***initial point and*** Q ***is the* terminal point *of*** $\vec{PQ}$.

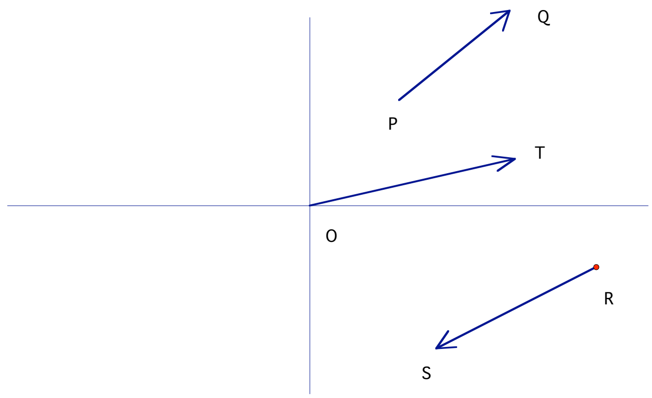


Figure 2.1.2: Examples of directed line segments

Example 2.1.1. Some examples in the plane are shown in the Figure (2.1.2).

Definition 2.2. A **directed line segment** $\vec{PQ}$ has a ***length***, which is defined to be $dist(P, Q)$, the distance of P to Q. We denote the length of $\vec{PQ}$ by $\parallel \vec{PQ} \parallel$. A directed line segment also has a direction - it defines a parallel class of lines as well as one of the two orientations on each of its lines.

Definition 2.3. We will say that two **directed line segments** $\vec{AB}$ and $\vec{CD}$ are ***equivalent*** provided they have the same **direction** and the same **length**. We will write $\vec{AB} \approx \vec{CD}$ to indicate that $\vec{AB}$ and $\vec{CD}$ are equivalent. By convention, any directed line segment $\vec{AB}$ will be equivalent to itself (so the **reflexive property** is satisfied).

Remark 2.1. When the **directed line segments** $\vec{AB}$ and $\vec{CD}$ are distinct the statement that $\vec{AB}$ and $\vec{CD}$ are **equivalent**, $\vec{AB} \approx \vec{CD}$, geometrically means that points A, B, C, D belong to a common plane and that $ABDC$ is a **parallelogram**. This follows from a theorem of Euclidean geometry which states: A quadrilateral $PQRS$ in the plane is a parallelogram if and only if some pair of opposite sides of the quadrilateral are parallel and equal in length.

Example 2.1.2. Let $A = (2, 5), B = (7, 3), C = (4, 8)$ and $D = (9, 6)$ then $\vec{AB} \approx \vec{CD}$. This is shown in Figure (2.1.3).

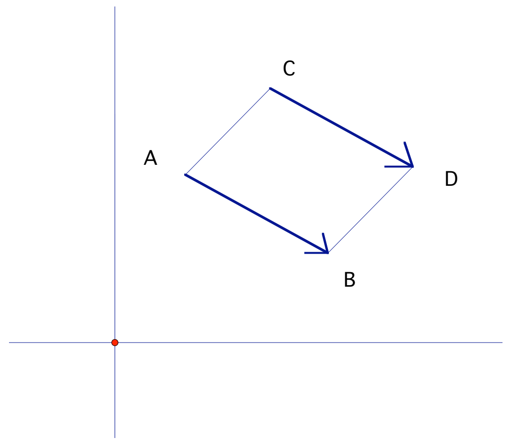


Figure 2.1.3: Equivalent line segments

As mentioned, we are assuming that the relation of being **equivalent** is **reflexive**. It is also clear that if $\vec{AB} \approx \vec{CD}$ then $\vec{CD} \approx \vec{AB}$ so that the relation $\approx$ is **symmetric** as well. Not quite as obvious is that the relation $\approx$ satisfies the **transtive property**. This is the subject of our first theorem:

Theorem 2.1.1. *If $\vec{AB} \approx \vec{CD}$ and $\vec{CD} \approx \vec{EF}$ then $\vec{AB} \approx \vec{EF}$. In words, if the* **directed line segment** *$\vec{AB}$ and $\vec{CD}$ are* **equivalent** *and $\vec{CD}$ is equivalent to $\vec{EF}$ then $\vec{AB}$ is equivalent to $\vec{EF}$.*

Proof. Since $\vec{AB} \approx \vec{CD}$ we know that $\| \vec{AB} \| = \| \vec{CD} \|$. Similarly, because $\vec{CD} \approx \vec{EF}$ we can conclude that $\| \vec{CD} \| = \| \vec{EF} \|$. Consequently, we can conclude that $\| \vec{AB} \| = \| \vec{EF} \|$

Also, since $\vec{AB} \approx \vec{CD}$, the **parallel class of lines** they define are same and similarly for $\vec{CD}$ and $\vec{EF}$. Moreover, since $\vec{AB}, \vec{CD}$ define the same **orientation** and $\vec{CD}, \vec{EF}$ likewise define the same orientation of this parallel class, this is true of $\vec{AB}$ and $\vec{EF}$ and therefore $\vec{AB}$ and $\vec{EF}$ have the same **direction**. Thus, $\vec{AB} \approx \vec{EF}$. This is illustrated in Figure (2.1.4). $\square$

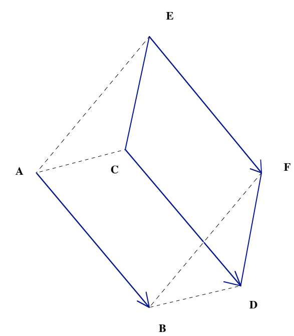


Figure 2.1.4: Transitivity of equivalence of directed line segments

For a **directed line segment** $\vec{AB}$ we denote by $[\vec{AB}]$ the collection of all directed line segments $\vec{CD}$ with $\vec{AB} \approx \vec{CD}$. This is the basis of our geometric interpretation of a vector:

Definition 2.4. By a ***geometric vector*** $\boldsymbol{v}$ we shall mean a collection $[\vec{AB}]$ of directed line segments. Any particular element of $\boldsymbol{v}$ is called a ***representative***.

In the next theorem we prove that if $\boldsymbol{v}$ is any vector then there exists a unique representative whose initial point is the origin.

Theorem 2.1.2. *Let $\boldsymbol{v}$ be a* **vector** *and X any point. Then there is a unique point Y such that $\boldsymbol{v} = [\vec{XY}]$. In particular, this holds for the origin $\mathbf{O}$; there is a unique point P such that $\boldsymbol{v} = [\vec{\mathbf{O}P}]$.*

Proof. Suppose our **vector** $\boldsymbol{v} = [\vec{AB}]$ is in 2-space, where $A = (a_1, a_2)$ and $B = (b_1, b_2)$. Set $P = (b_1 - a_1, b_2 - a_2)$. Then $\vec{AB} \approx \vec{\mathbf{O}P}$. More generally, if $X = (x_1, x_2)$ then set $Y = (x_1 + (b_1 - a_1), x_2 + (b_2 - a_2))$. $\vec{AB} \approx \vec{XY}$.

The proof for a vector in 3-space is done in exactly the same way.

□

Definition 2.5. The representative of the form $\vec{\mathbf{O}P}$ for a vector $\boldsymbol{v}$ is called the ***canonical representative of*** $\boldsymbol{v}$.

Remark 2.2. The **canonical representative** $\vec{\mathbf{O}P}$ of a vector $\boldsymbol{v}$ is uniquely determined by the **terminal point** P. This means we can identity this particular representative, and therefore the vector $\boldsymbol{v}$, by specifying just the terminal point P for the representative of $\boldsymbol{v}$ with initial point the origin, $\mathbf{O}$. To avoid confusing geometric vectors with points, when dealing with vectors we will write the coordinates vertically instead of horizontally.

Example 2.1.3. Let $A = (3,7), B = (5,11)$. Then $\vec{AB} \approx \vec{\mathbf{O}P}$ where $P = (2,4)$. We then write $\boldsymbol{v} = [\vec{AB}] = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$.

Example 2.1.4. Let $A = (2,7,11), B = (0,8,6)$. Then $\vec{AB} \approx \vec{\mathbf{O}P}$ where $P = (-2,1,-5)$. We write $\boldsymbol{v} = [\vec{AB}] = \begin{pmatrix} -2 \\ 1 \\ -5 \end{pmatrix}$.

The identification made in **Remark** (2.2) allows us to establish a correspondence between familiar geometric point sets, such as **lines** in 2-space and lines and **planes** in 3-space, with sets of vectors. However, in doing so we have to constantly remind ourselves that for a point set such as a plane in 3-space the corresponding vectors are not the vectors which lie in the plane (have their **initial and terminal points** in the plane) but those vectors with initial point the origin, $\mathbf{O}$, and having their terminal point in the plane. See Figure (2.1.5) and **Figure** (2.1.6).

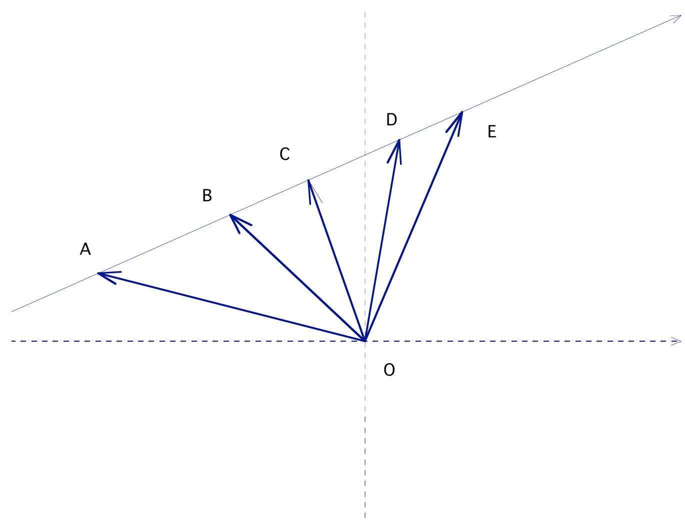


Figure 2.1.5: Correspondence between points on a line and a collection of vectors

The reason for formulating the **geometric notion of vector** in this way is that it allows us to move the line segments around "freely" and this provides us with a convenient way to define addition. Before doing so we introduce two important concepts: the zero vector and the negative of a vector.

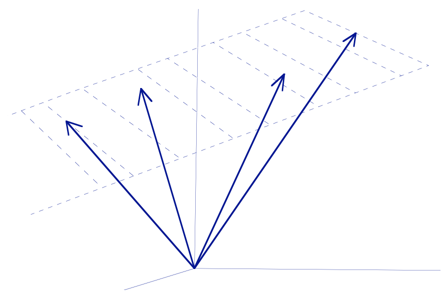


Figure 2.1.6: Correspondence between points on a plane and a collection of vectors

Definition 2.6. The ***geometric zero vector***, denoted by $\mathbf{0}$, is the vector $[\vec{AA}]$ where A is any point.

As we will see momentarily (after we define the sum of vectors) the **zero vector**, $\mathbf{0}$, has the property that, when is added to any vector $\boldsymbol{v}$ the result is $\boldsymbol{v}$, that is, it is unchanged. Thus, in this way the zero vector acts like the real number zero. We now define the negative of a vector $\boldsymbol{v}$.

Definition 2.7. For a vector $\boldsymbol{v} = [\vec{AB}]$ the ***negative*** of $\boldsymbol{v}$, denoted by $-\boldsymbol{v}$, is the vector $[\vec{BA}]$.

Remark 2.3. Let P be a point in 2- or 3-space and P^r its reflection in the origin. If $\boldsymbol{v} = [\vec{\mathbf{O}P}]$ then $-\boldsymbol{v} = [\vec{\mathbf{O}P^r}]$. This is illustrated in Figure (2.1.7).

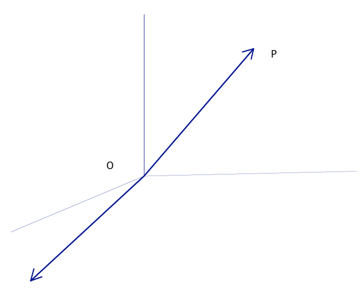


Figure 2.1.7: Negative of a vector

We now define how to add two vectors:

Definition 2.8. To ***add*** **vectors** $\boldsymbol{v}$ and $\boldsymbol{w}$ take any **representative** $\vec{AB}$ of $\boldsymbol{v}$ and $\vec{BC}$ of $\boldsymbol{w}$. Then $\boldsymbol{v}+\boldsymbol{w} = [\vec{AC}]$. We also say that $[\vec{AC}]$ is the ***sum*** of $\boldsymbol{v}$ and $\boldsymbol{w}$. See Figure (2.1.8).

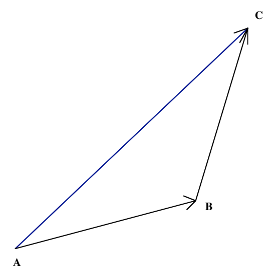


Figure 2.1.8: Sum of two vectors

Remark 2.4. This method for adding vectors is known as the ***parallelogram rule***. To see why this is so, let D be the unique point so that $\vec{AD} \approx \vec{BC}$. Then $ABCD$ is a parallelogram and this implies that $\vec{AB} \approx \vec{DC}$. The directed line segment $\vec{AC}$ is a diagonal of this parallelogram. See Figure (2.1.9).

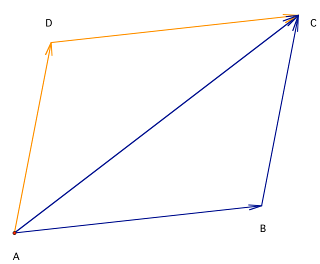


Figure 2.1.9: Parallelogram rule for adding vectors

Example 2.1.5. Let $\boldsymbol{v} = [\vec{\mathbf{O}P}] = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$ and $\boldsymbol{w} = [\vec{\mathbf{O}Q}] = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$. This means that $P = (3,7)$ and $Q = (2,-3)$. The directed line segment $\vec{PT}$ is in $\boldsymbol{w}$ where $T = (5,4)$. Thus, $\boldsymbol{v}+\boldsymbol{w} = [\vec{\mathbf{O}T}] = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$ This is shown in Figure (2.1.10).

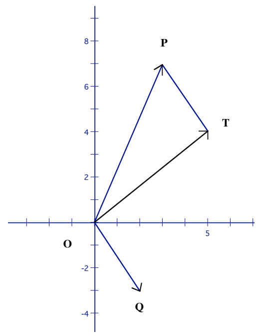


Figure 2.1.10: Example of adding two vectors

Remark 2.5. 1) We have not actually shown that definition of the **sum** of two **vectors** is "well defined." It could be that a choice of a different **representative** for $\boldsymbol{v}$ might result in a different outcome. To be completely rigorous we should prove that if $\vec{AB} \approx \vec{A'B'}$ and $\vec{BC} \approx \vec{B'C'}$ then $\vec{AC} \approx \vec{A'C'}$. See Figure (2.1.11). We leave this as a **challenge exercise**.

2) If $\boldsymbol{v} = \begin{pmatrix} a \\ b \end{pmatrix}$, that is, $\boldsymbol{v} = [\vec{\mathbf{O}P}]$ with $P = (a, b)$ and $\boldsymbol{w} = \begin{pmatrix} c \\ d \end{pmatrix} = [\vec{\mathbf{O}Q}]$ with $Q = (c, d)$ then $\boldsymbol{v} + \boldsymbol{w} = \begin{pmatrix} a+c \\ b+d \end{pmatrix}$. A similar formula holds for the sum of two vectors in 3-space.

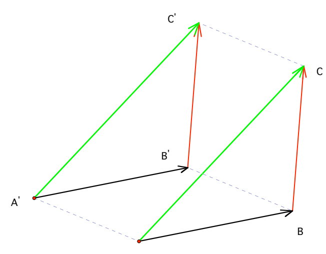


Figure 2.1.11: The sum of two vectors is well-defined

Another important concept is that of a (scalar) multiple of a **vector**.

Definition 2.9. Let $\boldsymbol{v}$ be a **vector** and c a positive real number. The ***scalar product*** of $\boldsymbol{v}$ by c, denoted by $c\boldsymbol{v}$, is the vector with the same **direction** as $\boldsymbol{v}$ but with **length** c times the length of $\boldsymbol{v}$. If c is a negative real number then we define $c\boldsymbol{v} = (-c)(-\boldsymbol{v})$. Finally, $0\boldsymbol{v} = \mathbf{0}$.

Remark 2.6. In 2-space, if $\boldsymbol{v} = [\vec{\mathbf{O}P}] = \begin{pmatrix} x \\ y \end{pmatrix}$ then $c\boldsymbol{v} = \begin{pmatrix} cx \\ cy \end{pmatrix}$. In 3-space, if $\boldsymbol{v} = [\vec{\mathbf{O}P}] = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ then $c\boldsymbol{v} = \begin{pmatrix} cx \\ cy \\ cz \end{pmatrix}$.

We collect some properties about **addition** and **scalar multiplication** of **vectors** and state these as the following theorem:

Theorem 2.1.3. *1. Addition of vectors is commutative: for any two vectors $\boldsymbol{v}, \boldsymbol{w}$ we have $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$.*

2. Addition of vectors is associative: for any three vectors $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}$ we have $(\boldsymbol{v}+\boldsymbol{w})+\boldsymbol{x} = \boldsymbol{v} + (\boldsymbol{w} + \boldsymbol{x})$.

3. Any vector $\boldsymbol{v}$ is unchanged when we add the zero vector to it, that is, $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$.

4. The sum of any vector $\boldsymbol{v}$ and its negative, $-\boldsymbol{v}$, is the zero vector, that is, $\boldsymbol{v}+(-\boldsymbol{v}) = \mathbf{0}$.

5. For any vector $\boldsymbol{v}$, $1\boldsymbol{v} = \boldsymbol{v}$ and $0\boldsymbol{v} = \mathbf{0}$.

6. For any scalar c and vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ we have $c(\boldsymbol{v} + \boldsymbol{w}) = c\boldsymbol{v} + c\boldsymbol{w}$.

7. For any scalars c and d and vector $\boldsymbol{v}$ we have $(c + d)\boldsymbol{v} = c\boldsymbol{v} + d\boldsymbol{v}$.

8. For any scalars c and d and vector $\boldsymbol{v}$ we have $d(c\boldsymbol{v}) = (dc)\boldsymbol{v}$.

Proof. These can all be proved geometrically. We illustrate with the proof of several and assign some of the others as **challenge exercises**. They are also treated in an algebraic fashion in the next section.

1) Let $\boldsymbol{v} = [\vec{\mathbf{O}P}]$ and $\boldsymbol{w} = [\vec{PQ}]$. Then $\boldsymbol{v} + \boldsymbol{w} = [\vec{\mathbf{O}Q}]$. On the other hand, let R be the unique point so that the quadrilateral $\mathbf{O}PQR$ is a **parallelogram**. Then $\boldsymbol{w} = [\vec{\mathbf{O}R}]$ and $\boldsymbol{w} = [\vec{RQ}]$. See Figure (2.1.12). Then $\boldsymbol{w} + \boldsymbol{v} = [\vec{\mathbf{O}Q}]$ as well.

2) Let $\boldsymbol{v} = [\vec{\mathbf{O}P}], \boldsymbol{w} = [\vec{PQ}], \boldsymbol{x} = [\vec{QR}]$. See Figure (2.1.13). Then $\boldsymbol{v} + \boldsymbol{w} = [\vec{\mathbf{O}Q}]$, and $(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{x} = [\vec{\mathbf{O}R}]$. On the other hand, $\boldsymbol{w} + \boldsymbol{x} = [\vec{QR}], \boldsymbol{v} + (\boldsymbol{w} + \boldsymbol{x}) = [\vec{\mathbf{O}R}]$. Therefore $(\boldsymbol{v} + \boldsymbol{w}) + \boldsymbol{x} = \boldsymbol{v} + (\boldsymbol{w} + \boldsymbol{x})$.

5) These follow immediately from the definition of **scalar multiplication**.

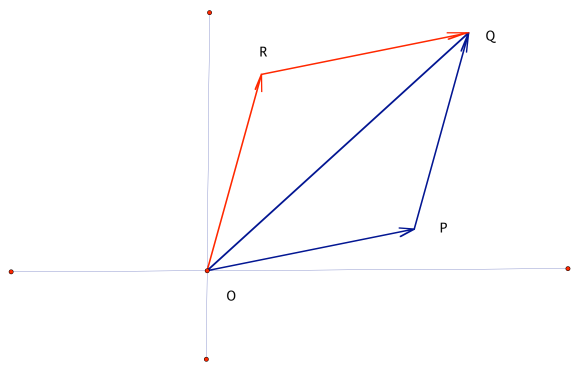


Figure 2.1.12: Addition of vectors is commutative

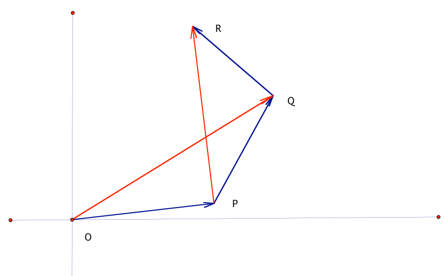


Figure 2.1.13: Addition of vectors is associative

6) If $c = 1$ then this is just part 5). Suppose $c > 1$. Let $\boldsymbol{v} = [\vec{\mathbf{O}P}], \boldsymbol{w} = [\vec{PQ}]$. Let $c\boldsymbol{v} = [\vec{\mathbf{O}P'}]$ and $c\boldsymbol{w} = [\vec{P'Q'}]$. See Figure (2.1.14).

Consider the two triangles $\triangle \mathbf{O}PQ$ and $\triangle \mathbf{O}P'Q'$. Side $\mathbf{O}P$ of $\triangle \mathbf{O}PQ$ lies on side $\mathbf{O}P'$ of $\triangle \mathbf{O}P'Q'$. Also, sides PQ and $P'Q'$ are **parallel**.

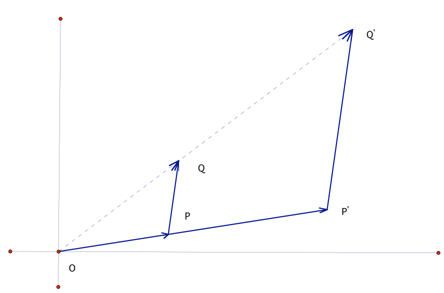


Figure 2.1.14: First distributive property for scalar multiplication

The ratios $\frac{\|[\vec{\mathbf{O}P'}]\|}{\|[\vec{\mathbf{O}P}]\|}$ and $\frac{\|[\vec{P'Q'}]\|}{\|[\vec{PQ}]\|}$ are both equal to c. This implies that the triangles are **similar**, that the point Q lies on the line segment joining $\mathbf{O}$ to Q' and that the ratio $\frac{\|[\vec{\mathbf{O}Q'}]\|}{\|[\vec{\mathbf{O}Q}]\|} = c$ as well. Consequently, $c\boldsymbol{v} + c\boldsymbol{w} = [\vec{\mathbf{O}Q'}] = c[\vec{\mathbf{O}Q}] = c(\boldsymbol{v} + \boldsymbol{w})$.

The case that $0 < c < 1$ is proved in a similar fashion.

Suppose now that $c = -1$. We must show that $(-\boldsymbol{v}) + (-\boldsymbol{w}) = -(\boldsymbol{v} + \boldsymbol{w})$.

Let $\boldsymbol{v} = [\vec{\mathbf{O}P}], \boldsymbol{w} = [\vec{PQ}], -\boldsymbol{v} = [\vec{\mathbf{O}P'}], -\boldsymbol{w} = [\vec{P'Q'}]$. $\boldsymbol{v} + \boldsymbol{w} = [\vec{\mathbf{O}Q}]$ and $(-\boldsymbol{v}) + (-\boldsymbol{w}) = [\vec{\mathbf{O}Q'}]$. See Figure (2.1.15).

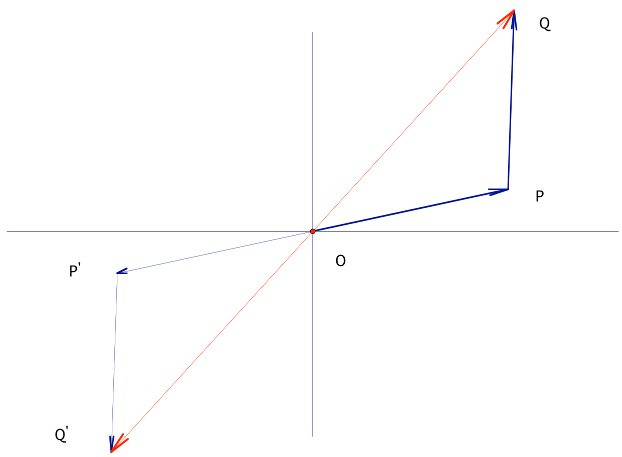


Figure 2.1.15: Taking negatives distributes over addition

Consider the triangles $\triangle\mathbf{O}PQ$ and $\triangle\mathbf{O}P'Q'$. The **length** of side $\mathbf{O}P$ is equal to the length of side $\mathbf{O}P'$ and the length of side PQ is equal to the length of side $P'Q'$. Also the line of segment PQ is parallel to the line of segment $P'Q'$. This implies that $\angle\mathbf{O}PQ$ is **congruent** to $\angle\mathbf{O}P'Q'$ and consequently the triangles are **congruent triangles**.

Since the triangles are congruent, the length of side $\mathbf{O}Q$ is equal to the length of side $\mathbf{O}Q'$. It also implies that $\angle P\mathbf{O}Q$ is congruent to $\angle P'\mathbf{O}Q'$ and $\angle\mathbf{O}QP$ is congruent to $\angle\mathbf{O}Q'P'$. Now the **measure** of $\angle Q'\mathbf{O}P$ is equal to the sum of the measures of $\angle\mathbf{O}PQ$ and $\angle\mathbf{O}QP$ and the measure of $\angle P'\mathbf{O}Q$ is equal to the sum of the measures of $\angle\mathbf{O}P'Q'$ and $\angle\mathbf{O}Q'P'$.

Therefore $\angle Q'\mathbf{O}P$ and $\angle P'\mathbf{O}Q$ have the same **measure** and are therefore congruent. Since $\angle P\mathbf{OQ}$ and $\angle P'\mathbf{O}Q'$ are congruent, it follows that $\angle Q'\mathbf{O}Q$ and $\angle P'\mathbf{O}P$ are congruent and therefore are straight angles. Since the length of $\mathbf{O}Q$ and $\mathbf{O}Q'$ are equal this implies that $[\vec{\mathbf{O}Q'}] = -[\vec{\mathbf{O}Q}]$. This proves that $-(\boldsymbol{v} + \boldsymbol{w}) = (-\boldsymbol{v}) + (-\boldsymbol{w})$.

Now assume that $c < 0$. By definition of the scalar product $c(\boldsymbol{v}+\boldsymbol{w}) = (-c)[-(\boldsymbol{v}+\boldsymbol{w})]$. By what we proved immediately above,

$$(-c)[-(\boldsymbol{v}+\boldsymbol{w})] = (-c)[(-\boldsymbol{v}) + (-\boldsymbol{w})] \tag{2.1}$$

Since $-c > 0$ we can apply the case for positive scalars to conclude from (2.1) that

$$(-c)[(-\boldsymbol{v}) + (-\boldsymbol{w})] = (-c)(-\boldsymbol{v}) + (-c)(-\boldsymbol{w}) = c\boldsymbol{v} + c\boldsymbol{w} \tag{2.2}$$

by the definition of scalar product for negative c.

We leave 3) and 4) as **challenge exercises** and will prove the others algebraically in the next section. □

A natural question to ask is whether we can describe geometrically what is obtained when we combine **vectors** using **addition** and **scalar multiplication** . The answer is yes and the objects that arise are quite familiar: **lines** and **planes** which pass through the origin. Other lines and planes are realized as “translates” of lines and planes which pass through the origin.

Before we do some examples, first a reminder about the correspondence between geometric point sets and collections of vectors: Given a point set such as a **plane** or a **line**, in general the corresponding collection of vectors **is not** the collection of all directed line segments on the line or plane. Rather it is the collection of all vectors with a representative $\vec{\mathbf{O}P}$ with **initial point** the origin $\mathbf{O}$ and **terminal point** P lying on the point set (line, plane, etc.).

Example 2.1.6. Consider the single **vector** $\boldsymbol{v}$ with **canonical representative** $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$. What do we get when we take all **scalar multiples** of $\boldsymbol{v}$?

We get all vectors $[\vec{\mathbf{O}Q}]$ which have canonical representative $\begin{pmatrix} t \\ 2t \end{pmatrix}$ for some real number t. If $Q = (x, y)$ then $y = 2x$ or $2x - y = 0$, the line through the origin with slope 2. We will say that the vector $\boldsymbol{v}$ ***spans*** this line. See Figure (2.1.16).

Example 2.1.7. Let $\boldsymbol{v}$ be the **vector** in 3-space with **canonical representative** $\begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$.

A typical **scalar multiple** $t\boldsymbol{v}$ has canonical representative $\begin{pmatrix} 2t \\ -3t \\ t \end{pmatrix}$. The point of 3-space corresponding to $t\boldsymbol{v}$ is $Q = (2t, -3t, t)$. The coordinates of such a point satisfy the parametric equations

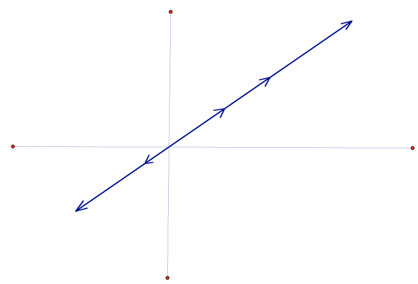


Figure 2.1.16: Span of a single vector

$$x = 2t, y = -3t, z = t$$

The collection $\mathcal{L}$ of these points is a line through the origin.

Definition 2.10. When a line $\mathcal{L}$ through the origin arises as all multiples of a **vector** $\boldsymbol{v}$ we say the vector $\boldsymbol{v}$ ***spans*** the line $\mathcal{L}$ or that the ***span*** of $\boldsymbol{v}$ is $\mathcal{L}$.

Example 2.1.8. Now suppose we have two **vectors**, $\boldsymbol{v}$ and $\boldsymbol{w}$, in 3-space which are not multiples of each other. What do we get when we **add** arbitrary multiples of $\boldsymbol{v}$ and $\boldsymbol{w}$?

We saw in **Example** (2.1.7) the multiples of $\boldsymbol{v}$ span a line $\mathcal{L}$ through the origin and likewise the multiples of $\boldsymbol{w}$ span a line $\mathcal{M}$ through the origin which is distinct from $\mathcal{L}$ since we are assuming that $\boldsymbol{v}$ and $\boldsymbol{w}$ are not multiples.

These two lines determine a plane Π through the origin. Adding any vector $[\vec{\mathbf{O}P}]$ with P on $\mathcal{L}$ and any vector $[\vec{\mathbf{O}Q}]$ with Q on $\mathcal{M}$ we get a vector $[\vec{\mathbf{O}R}]$ with the point R in the plane Π. Moreover, any vector $[\vec{\mathbf{O}R}]$ with the point R in the plane Π can be obtained as a sum in this way. We say the two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ ***span the plane*** Π. See Figure (2.1.17).

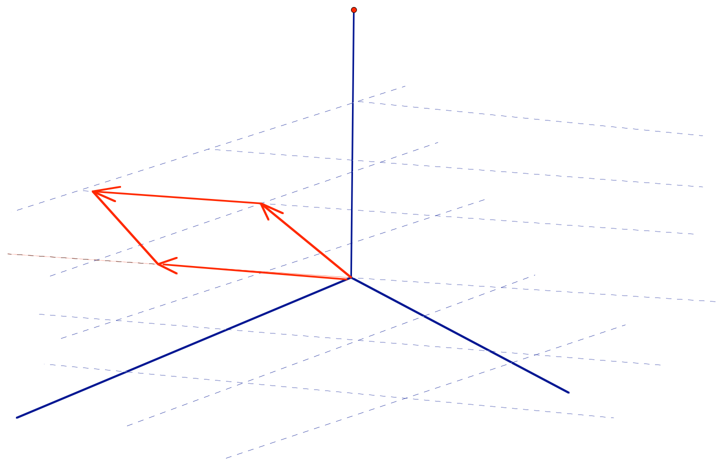


Figure 2.1.17: Span of two vectors in three space

More concretely, suppose $\boldsymbol{v}$ has **canonical representative** $\begin{pmatrix}1\\1\\1\end{pmatrix}$ and $\boldsymbol{w}$ has canonical representative $\begin{pmatrix}1\\-1\\0\end{pmatrix}$. The span of $\boldsymbol{v}$ consists of all **vectors** $[\vec{\mathbf{O}P}]$ where P is on the line $\mathcal{L}$ with parametric equations

$$x = s, y = s, z = s$$

The span of $\boldsymbol{w}$ consists of all vectors $[\vec{\mathbf{O}Q}]$ where Q is on the line $\mathcal{M}$ with parametric equations

$$x = t, y = -t, z = 0$$

A **sum** of a multiple of $\boldsymbol{v}$ and a multiple of $\boldsymbol{w}$, $s\boldsymbol{v} + t\boldsymbol{w}$, has canonical represenative $\vec{\mathbf{O}R}$ where $R = (s+t, s-t, s)$. Notice that each such point satisfies the equation $x + y - 2z = 0$ since $(s+t) + (s-t) - 2s = 0$.

So all possible sums of a multiple of $\boldsymbol{v}$ and a multiple of $\boldsymbol{w}$ is contained in the plane $\Pi = \{(x, y, z) : x + y - 2z = 0\}$ which passes through the origin.

On the other hand, we claim that any vector $[\vec{\mathbf{O}R}]$ where R is on the plane Π is a sum of a multiple of $\boldsymbol{v}$ and a multiple of $\boldsymbol{w}$.

For example, suppose $R = (x, y, z)$ satisfies $x + y - 2z = 0$. Set $s = \frac{x+y}{2}, t = \frac{x-y}{2}$. Since $x + y - 2z = 0, z = \frac{x+y}{2} = s$. Then $[\vec{\mathbf{O}R}] = s\boldsymbol{v} + t\boldsymbol{w}$.

Definition 2.11. Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be two vectors. The collection of all possible sums of a multiple $s\boldsymbol{v}$ of $\boldsymbol{v}$ and $t\boldsymbol{w}$ of $\boldsymbol{w}$ is called the ***span*** of $\boldsymbol{v}$ and $\boldsymbol{w}$.

The notion of the **span** of **vectors** can be extended to three or more vectors. In the plane, two vectors which are not multiples of one other will span all possible vectors. In 3-space, as we have seen in **Example** (2.1.8), two vectors which are not multiples span a plane Π. If we were to take a third vector $\boldsymbol{x} = [\vec{\mathbf{O}S}]$ where S does not lie in Π, then every vector in 3-space can be obtained as a sum of a multiple of $\boldsymbol{v}, \boldsymbol{w}$ and $\boldsymbol{x}$.

What about other **lines** or **planes** - those that do not pass through the origin. We illustrate with an example.

Example 2.1.9. Consider the line $\mathcal{L}$ consisting of those points P with coordinates (x, y) which satisfy

$$2x - y = 8 \tag{2.3}$$

This line is parallel to the line $\mathcal{M}$ which passes through the origin and whose points satisfy the equation $2x - y = 0$. We saw in **Example** (2.1.6) that $\mathcal{M}$ may be identified with the span of the vector $\boldsymbol{v}$ which has canonical form $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

The point $P = (4, 0)$ lies on $\mathcal{L}$ and so $\boldsymbol{p} = [\vec{\mathbf{O}P}]$ is a particular example of a vector identified with a point on $\mathcal{L}$.

Any other vector $\boldsymbol{u} = [\vec{\mathbf{O}Q}]$ with a point Q on $\mathcal{L}$ can be expressed as a sum of $\boldsymbol{p}$ and a vector $\boldsymbol{x}$ identified with a point of $\mathcal{M}$.

For example, let $Q = (7, 6)$, a point on $\mathcal{L}$ and set $R = (3, 6)$, a point on $\mathcal{M}$. Then we have the following (see Figure (2.1.18))

$$[\vec{\mathbf{O}Q}] = [\vec{\mathbf{O}P}] + [\vec{\mathbf{O}R}] \tag{2.4}$$

In general, the collection of vectors which are identified with $\mathcal{L}$ can be expressed as the sum of the particular vector $\boldsymbol{p}$ and a vector identified with a point on the line $\mathcal{M}$. We saw in Example (2.1.6) that the **vectors** identified with $\mathcal{M}$ are just the **scalar multiples** of the $\boldsymbol{v}$. Therefore, we can write

$$\boldsymbol{u} = \boldsymbol{p} + t\boldsymbol{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{2.5}$$

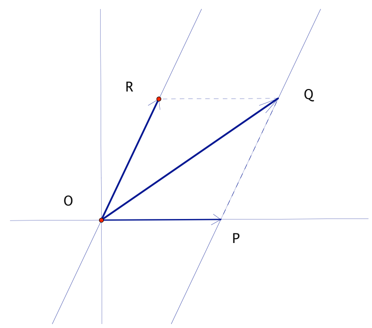


Figure 2.1.18: The set of vectors corresponding to the points on an arbitrary line

Example 2.1.10. Consider the plane Ω consisting of those points P with coordinates (x, y, z) which satisfy the equation

$$x + y - 2z = -6 \tag{2.6}$$

This plane is parallel to the plane Π which passes through the origin and whose points satisfy the equation $x + y - 2z = 0$ which was the subject of **Example** (2.1.8).

Let $P = (0, 0, 3)$. Then $\boldsymbol{p} = [\vec{\mathbf{O}P}]$ is an example of a particular **vector** identified with a point on the plane Ω.

Any other vector $\boldsymbol{u} = [\vec{\mathbf{O}Q}]$, with a point Q on Ω, can be expressed as a sum of $\boldsymbol{p}$ and a vector $\boldsymbol{x}$ identified with a point on Π.

For example, let $Q = (2, -2, 3)$, a point on Ω, and set $R = (2, -2, 0)$, a point on Π. Then we have $[\vec{\mathbf{O}Q}] = [\vec{\mathbf{O}P}] + [\vec{\mathbf{O}R}]$ (see Figure (2.1.19)).

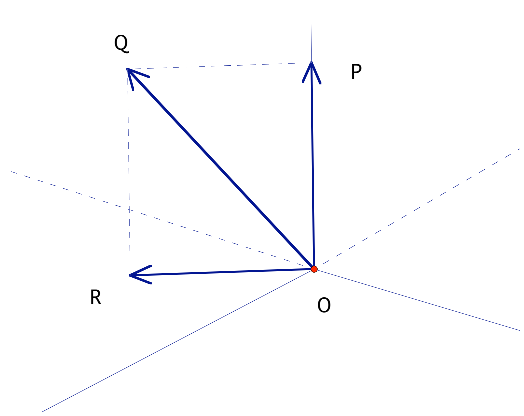


Figure 2.1.19: Vectors corresponding to points on an arbitrary plane in three space

In general, the collection of **vectors** which are identified with Ω can be expressed as the **sum** of the particular vector $\boldsymbol{p}$ with a vector identified with the plane Π. We previously saw in **Example** (2.1.8) that Π consists of the **span** of the vectors $\boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{w} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, that is, all possible expressions of the form $s\boldsymbol{v} + t\boldsymbol{w}$. Thus, for an arbitrary vector $\boldsymbol{u}$ on Ω there exists scalars s and t such that

$$\boldsymbol{u} = \boldsymbol{p} + s\boldsymbol{v} + t\boldsymbol{w} \tag{2.7}$$

In terms of canonical forms this can be written as

$$\boldsymbol{u} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} + s\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3+s+t \\ s-t \\ s \end{pmatrix} \tag{2.8}$$

What You Can Now Do

1. Given a **vector** $\boldsymbol{v} = [\vec{AB}]$, find the length of $\boldsymbol{v}$, $\parallel \boldsymbol{v} \parallel$.

2. Given a **vector** $\boldsymbol{v} = [\vec{AB}]$, find the **canonical representative** of $\boldsymbol{v}$.

3. Given a **vector** $\boldsymbol{v} = [\vec{AB}]$ and a point C, find a point D such that $\boldsymbol{v} = [\vec{CD}]$.

4. Given **vectors** $\boldsymbol{v} = [\vec{AB}]$ and $\boldsymbol{w} = [\vec{CD}]$, compute the **sum** of the vectors using **canonical representatives**.

5. Given a **vector** $\boldsymbol{v} = [\vec{AB}]$ and a scalar c, compute the **scalar product** $c\boldsymbol{v}$ and determine its **canonical representative**.

Method (How to do it)

Method 2.1.1. Given a vector $\boldsymbol{v} = [\vec{AB}]$ find the **length** of $\boldsymbol{v}$, $\parallel \boldsymbol{v} \parallel$.

Use the distance formula. If $A = (a_1, a_2), B = (b_1, b_2)$ are in the plane, then

$$\parallel [\vec{AB}] \parallel = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

$A = (a_1, a_2, a_3), B = (b_1, b_2, b_3)$ are points in 3-space then

$$\parallel [\vec{AB}] \parallel = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2}.$$

Example 2.1.11. Find the length of the vector $[\vec{AB}]$ where $A = (4, 9), B = (7, 5)$.

$$\| [\vec{AB}] \| = \sqrt{(7-4)^2 + (5-9)^2} = \sqrt{3^2 + (-4)^2} = \sqrt{9+16} = \sqrt{25} = 5.$$

Example 2.1.12. Find the length of the vector $[\vec{\mathbf{O}P}]$ where $P = (4, -8)$.

$$\| [\vec{\mathbf{O}P}] \| = \sqrt{4^2 + (-8)^2} = \sqrt{16+64} = \sqrt{80} = 4\sqrt{5}$$

Example 2.1.13. Find the length of the vector $[\vec{AB}]$ where $A = (2, 5, 9), B = (-4, 5, 13)$.

$$\| [\vec{AB}] \| = \sqrt{[(-4)-2]^2 + [5-5]^2 + [13-9]^2} =$$
$$\sqrt{(-6)^2 + 0^2 + 4^2} = \sqrt{36+16} = \sqrt{52} = 2\sqrt{13}.$$

Example 2.1.14. Find the length of the vector $[\vec{\mathbf{O}P}]$ where $P = (-2, 1, 2)$.

$$\| [\vec{\mathbf{O}P}] \| = \sqrt{(-2)^2 + 1^2 + 2^2} = \sqrt{9} = 3.$$

Method 2.1.2. 2. Given a vector $\boldsymbol{v} = [\vec{AB}]$, find the canonical representative of $\boldsymbol{v}$.

Suppose $A = (a_1, a_2), B = (b_1, b_2)$ are in the plane. Then the canonical representative of $\boldsymbol{v}$ is $\begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \end{pmatrix}$.

If $A = (a_1, a_2, a_3), B = (b_1, b_2, b_3)$ are in 3-space then the canonical representative of $\boldsymbol{v}$ is $\begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \\ b_3 - a_3 \end{pmatrix}$.

Example 2.1.15. Let $\boldsymbol{v} = [\vec{AB}]$ where $A = (2, 6), B = (7, -1)$. Find the canonical representative of $\boldsymbol{v}$.

The canonical representative of $\boldsymbol{v}$ is $\begin{pmatrix} 7-2 \\ -1-6 \end{pmatrix} = \begin{pmatrix} 5 \\ -7 \end{pmatrix}$

Example 2.1.16. Let $\boldsymbol{v} = [\vec{AB}]$ where $A = (2, -5, 8), B = (-3, 0, -2)$. Find the canonical representative of $\boldsymbol{v}$.

The canonical representative of $\boldsymbol{v}$ is $\begin{pmatrix} -3-2 \\ 0-(-5) \\ -2-8 \end{pmatrix} = \begin{pmatrix} -5 \\ 5 \\ -10 \end{pmatrix}$

Method 2.1.3. Given a **vector** $\boldsymbol{v} = [\vec{AB}]$ and a point C, find a point D such that $\boldsymbol{v} = [\vec{CD}]$.

If $C = (c_1, c_2)$ is in the plane, and the **canonical representative** of $\boldsymbol{v}$ is $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ then $D = (c_1 + v_1, c_2 + v_2)$. Therefore, if $\boldsymbol{v} = [\vec{AB}]$ with A, B as above, then $D = (c_1 + (b_1 - a_1), c_2 + (b_2 - a_2))$.

If $C = (c_1, c_2, c_3)$ is in 3-space and the **canonical representative** of $\boldsymbol{v}$ is $\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ then $D = (c_1 + v_1, c_2 + v_2, c_3 + v_3)$. Therefore, if $\boldsymbol{v} = [\vec{AB}]$ with A, B as above, then $D = (c_1 + (b_1 - a_1), c_2 + (b_2 - a_2), c_3 + (b_3 - a_3))$.

Example 2.1.17. If $A = (-2, 2), B = (3, 0), \boldsymbol{v} = [\vec{AB}]$, and $C = (4, -2)$ find D such that $\boldsymbol{v} = [\vec{CD}]$.

The **canonical representative** of $\boldsymbol{v}$ is $\begin{pmatrix} 5 \\ -2 \end{pmatrix}$. Then $D = (4 + 5, -2 + (-2)) = (9, -4)$.

Example 2.1.18. If $A = (1, 2, -3), B = (-2, 3, -4), \boldsymbol{v} = [\vec{AB}]$, and $C = (4, 3, 2)$ find D such that $\boldsymbol{v} = [\vec{CD}]$.

The **canonical representative** of $\boldsymbol{v}$ is $\begin{pmatrix} -3 \\ 1 \\ -1 \end{pmatrix}$ Then $D = (4+(-3), 3+1, 2+(-1)) = (1, 4, 1)$.

Method 2.1.4. Given **vectors** $\boldsymbol{v} = [\vec{AB}]$ and $\boldsymbol{w} = [\vec{CD}]$, compute the **sum** of the vectors using **canonical representatives**.

Use **Method** (2.1.2) to find the **canonical representatives** of $\boldsymbol{v}$ and $\boldsymbol{w}$.

If $\boldsymbol{v}$ has **canonical representative** $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ and $\boldsymbol{w}$ has canonical representative $\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ then the canonical representative of $\boldsymbol{v} + \boldsymbol{w}$ is $\begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix}$

If $\boldsymbol{v}$ has **canonical representative** $\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ and $\boldsymbol{w}$ has canonical representative $\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$ then $\boldsymbol{v} + \boldsymbol{w}$ has canonical representative $\begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ v_3 + \boldsymbol{w}_3 \end{pmatrix}$.

Example 2.1.19. Let $\boldsymbol{v} = [\vec{AB}]$ with $A = (1, -3, 7), B = (3, 6, 2)$ and $\boldsymbol{w} = [\vec{CD}]$ with $C = (2, -4, 8)$ and $D = (1, 1, 2)$.

Compute $\boldsymbol{v} + \boldsymbol{w}$ using **canonical representatives**.

The **canonical representative** of $\boldsymbol{v}$ is $\begin{pmatrix} 3-1 \\ 6-(-3) \\ 2-7 \end{pmatrix} = \begin{pmatrix} 2 \\ 9 \\ -5 \end{pmatrix}$. The canonical representative of $\boldsymbol{w}$ is $\begin{pmatrix} 1-2 \\ 1-(-4) \\ 2-8 \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \\ -6 \end{pmatrix}$.

Then the **canonical representative** of $\boldsymbol{v} + \boldsymbol{w}$ is

$$\begin{pmatrix} 2+(-1) \\ 9+5 \\ -5+(-6) \end{pmatrix} = \begin{pmatrix} 1 \\ 14 \\ -11 \end{pmatrix}$$

Method 2.1.5. Given a **vector** $\boldsymbol{v} = [\vec{AB}]$ and a scalar c compute the **canonical representative** of the **scalar product**, $\boldsymbol{v}$. Find a point C such that $c\boldsymbol{v} = [\vec{AC}]$.

Use **Method** (2.1.2) to find the **canonical representative**. This will be of the form $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ or $\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ depending on whether the vector is in the plane or 3-space. Then multiply each v_i by c. For the second part, use **Method** (2.1.3).

Example 2.1.20. Let $\boldsymbol{v} = [\vec{AB}]$ where $A = (2, -3, 5)$ and $B = (4, 3, -9)$. For each of the following scalars c compute the canonical representative of $c\boldsymbol{v}$. In each instance find a point C such that $c\boldsymbol{v} = [\vec{AC}]$.

a) $c = 2$
b) $c = \frac{1}{2}$
c) $c = -1$
d) $c = -3$.

The **canonical representative** of $\boldsymbol{v}$ is $\begin{pmatrix} 4-2 \\ 3-(-3) \\ -9-5 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ -14 \end{pmatrix}$.

a) The canonical representative of $2\boldsymbol{v}$ is $\begin{pmatrix} 4 \\ 12 \\ -28 \end{pmatrix}$. We find C by adding these components to the coefficients of $A : C = (2+4, -3+12, 5+(-28)) = (6, 9, -23)$.

b) The **canonical representative** of $\frac{1}{2}\boldsymbol{v}$ is $\begin{pmatrix} 1 \\ 3 \\ -7 \end{pmatrix}$. Then $C = (2+1, -3+3, 5+(-7)) = (3, 0, -2)$.

c) The **canonical representative** of $-\boldsymbol{v}$ is $\begin{pmatrix} -2 \\ -6 \\ 14 \end{pmatrix}$. In this case, $C = (2+(-2), -3+(-6), 5+14) = (0, -9, 19)$.

d) The **canonical representative** of $(-3)\boldsymbol{v}$ is $\begin{pmatrix} -6 \\ -18 \\ 42 \end{pmatrix}$. Now, $C = (2+(-6), -3+(-18), 5+42) = (-4, -21, 47)$.

Exercises

In exercises 1 - 6 find the **length of the vector** $[\vec{AB}]$ for the given points A and B. See **Method** (2.1.1).

1. $A = (2, 7), B = (8, -1)$

2. $A = (-2, 5), B = (4, -1)$

3. $A = (-3, -5), B = (3, 5)$

4. $A = (2, 5, -8), B = (4, -3, -7)$

5. $A = (-2, 4, -1), B = (10, 1, 3)$

6, $A = (8, -4, 3), B = (5, -7, 0)$

In exercises 7 - 9 let $A = (4, 7)$ and $C = (-2, 6)$. For each point B find the **canonical representative** of $[\vec{AB}]$ and a point D such that $\vec{AB} \approx \vec{CD}$. See **Method** (2.1.2) and **Method** (2.1.3).

7. $B = (2, 11)$

8. $B = (-1, 2)$

9. $B = (7, 11)$

In exercises 10 - 12 let $A = (-3, 1, 5)$ and $C = (-1, 5, 11)$. For each point B find the **canonical representative** of $[\vec{AB}]$ and a point D such that $\vec{AB} \approx \vec{CD}$. See **Method** (2.1.2) and **Method** (2.1.3).

10. $B = (-2, 3, 8)$

11. $B = (-7, -7, -6)$

12, $B = (1, 5, 9)$

In exercises 13 - 16 for the given vectors $\boldsymbol{v} = [\vec{AB}]$ and $\boldsymbol{w} = [\vec{CD}]$ compute the **sum** using **canonical representatives**. See **Method** (2.1.4).

13. $A = (1, 3), B = (4, -1), C = (7, -5), D = (10, -9)$

14. $A = (-3, 4), B = (1, 8), C = (2, -3), D = (3, 1)$

15. $A = (1, 2, 3), B = (3, 4, 1), C = (1, 2, 3), D = (-1, 0, 5)$

16. $A = (-1, 3, 8), B = (2, 6, 2), C = (3, 4, 5), D = (2, 3, 7)$

In exercises 17-24 you are given a vector $\boldsymbol{v} = [\vec{AB}]$ and a scalar c. Compute the **canonical representative** of $c\boldsymbol{v}$ and in each instance find a point C such that $c\boldsymbol{v} = [\vec{AC}]$. See **Method** (2.1.5).

17. $A = (2, 5), B = (-3, 7), c = 2.$

18. $A = (4, -1), B = (-2, 8), c = \frac{1}{3}.$

19. $A = (0, 4), B = (-2, 1), c = -2.$

20. $A = (3, 3), B = (7, -3), c = -\frac{1}{2}$

21. $A = (1, 2, 4), B = (3, 6, 12), c = -1.$

22. $A = (1, 2, 4), B = (3, 6, 12), c = \frac{3}{2}.$

23. $A = (4, -3, 7), B = (5, -1, 10), c = -2.$

24. $A = (4, 0, 6), B = (7, 6, -3), c = -\frac{1}{3}$

Challenge Exercises (Problems)

1. Assume that $\vec{AB} \approx \vec{A'B'}, \vec{BC} \approx \vec{B'C'}$. Prove that $\vec{AC} \approx \vec{A'C'}$.

2. Prove part 3) of Theorem (2.1.3).

3. Prove part 4) of Theorem (2.1.3).

Quiz Solutions

1. $\sqrt{(2-[-2])^2 + (7-4)^2} = \sqrt{4^2 + 3^2} =$

$\sqrt{16+9} = \sqrt{25} = 5.$

2. $\sqrt{(5-[-3])^2 + (4-0)^2 + (1-2)^2} = \sqrt{8^2 + 4^2 + 1^2} =$

$\sqrt{64+16+1} = \sqrt{81} = 9.$

3. $S = (2, 1).$

2.2. Vectors and the Space $\mathbb{R}^n$

Motivated by the previous section we introduce the general notion of an n-vector as well as the concept of $n-$space, denoted by $\mathbb{R}^n$. We develop the algebra of vectors and show how to represent a linear system as a vector equation and translate questions about a linear system into questions about vectors and the $n-$space $\mathbb{R}^n$.

Am I Ready for This Material

Readiness Quiz

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

The following are previous concepts and procedures that you need to be familiar with in this section:

linear system

solution of a linear system

solution set of a linear system

consistent linear system

inconsistent linear system

equivalent linear systems

echelon form of a linear system

leading variable

free variable

matrix

augmented matrix of a linear system

row equivalence of matrices

matrix in row echelon form

matrix in reduced row echelon form

echelon form of a matrix

reduced echelon from of a matrix

pivot positions of a matrix

pivot columns of a matrix

directed line segment

length of a directed line segment

equivalence of directed line segments

geometric vector

canonical representative of a geometric vector

geometric definition of zero vector

geometric definition of the negative of a vector

geometric sum of vectors

scalar multiple of a geometric vector

span of two geometric vectors

The following are algorithms that we will use in this section:

Gaussian elimination

procedure for solving a linear system

Quiz

1. Find the **reduced echelon form** of the matrix below.

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

2. A is a 4×4 matrix and **row equivalent** to the matrix

$$\begin{pmatrix} 1 & 3 & -1 & 2 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & -3 & -9 \\ 1 & 3 & 2 & 5 \end{pmatrix}$$

What are the **pivot columns** of A?

The **augmented matrix** of a **linear system** is $[A|\bar{b}]$ and this matrix is **row equivalent** to

$$\begin{pmatrix} 1 & -2 & 0 & -1 & | & 4 \\ 0 & 0 & 1 & 2 & | & -2 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$$

3. The number of **solutions** to this **linear system** is

a) Zero, the system is **inconsistent**
b) One, there is a unique solution.
c) There are infinitely many solutions
d) It is impossible to determine the number of solutions from the given information.

4. Which of the following are **solutions** of the **linear system** described above.

a) $\begin{pmatrix} 4 \\ 0 \\ -2 \\ 0 \end{pmatrix}$ b) $\begin{pmatrix} 0 \\ -2 \\ -2 \\ 0 \end{pmatrix}$ c) $\begin{pmatrix} 3 \\ 0 \\ 0 \\ -1 \end{pmatrix}$ d) $\begin{pmatrix} 4 \\ 1 \\ 2 \\ -2 \end{pmatrix}$ e) $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

Solutions

New Concepts

This is an extremely important section in which several fundamental concepts are introduced which will be used throughout the text. These are:

an $n-$vector

equality of $n-$vectors

$n-$space, $\mathbb{R}^n$

adddition of $n-$vectors

scalar multiplication of $n-$vectors

negative of a vector

the zero vector

the standard basis for $\mathbb{R}^n$

linear combination of vectors

Theory (Why It Works)

In Section (2.1) we introduced a notion of **vectors** in two and three diimensional **Euclidean space**. However, we do not want to limit ourselves to these spaces. We further saw that to each vector in 2-space (respectively, 3-space) there corresponds an ordered pair (respectively, ordered 3-tuple) of numbers. Moreover, the operations we introduced, **addition** and **scalar multiplication**, can be realized by operating on these ordered pairs, respectively, ordered triples. This gives us a way to generalize the notion of vector and to define what we mean by $n-$space: take a vector to be an $n-$tuple of numbers. This is captured in the following definition.

Definition 2.12. By a ***vector*** we will mean a single column of numbers. If the column has n entries, that is, it is an $n \times 1$ matrix, then we will refer to it as an $n-$vector. The entries which appear in a vector are called its ***components***.

Some examples are

$$\boldsymbol{u} = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 0 \\ -1 \\ 4 \\ 2 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ 0 \end{pmatrix}.$$

$\boldsymbol{u}$ is a 3-vector, $\boldsymbol{v}$ is a 4-vector and $\boldsymbol{w}$ is a 5-vector.

Example 2.2.1. Though motivated by the previous section,**vector**, as defined here, need not have any connection to geometry. The following are examples:

1) The 111-vector with **components** $\boldsymbol{p}^{Ca} = \begin{pmatrix} p_0^{Ca} \\ p_1^{Ca} \\ \vdots \\ p_{110}^{Ca} \end{pmatrix}$ records the number of persons living in California of age $i = 0, 1, 2, \ldots, 110$. If P^{Ca} is the total population of California and $d_i^{Ca} = \frac{p_i^{Ca}}{P^{Ca}}$ then the vector $\begin{pmatrix} d_0^{Ca} \\ d_1^{Ca} \\ \vdots \\ d_{110}^{Ca} \end{pmatrix}$ gives the percentage of the California population of age $i = 0, 1, \ldots, 110$.

2) The **4-vector** $\boldsymbol{v}_{PC} = \begin{pmatrix} 2.5 \\ 18.5 \\ 3 \\ 110 \end{pmatrix}$ records grams of protein, carbohydrates, fat and calories in a one ounce serving of Maple Pecan clusters. The vector $\boldsymbol{v}_{OF} = \begin{pmatrix} 3 \\ 23 \\ 1 \\ 110 \end{pmatrix}$ records the same information for a one ounce serving of Trader Joe's Oak Flakes cereal.

3) A chemical is used in the production of fertilizer, which is made from nitrogen, phosphoric acid and potash. We can associate with this chemical a **3-vector** indicating the percentage, by weight, of this chemical made up of these precursors to the desired

fertilizer. For example, the vector $\begin{pmatrix} 0.25 \\ 0.05 \\ 0.05 \end{pmatrix}$ indicates that this particular chemical is 25% nitrogen, 5% phosphoric acid and 5% potash.

Definition 2.13. By **n-space** we will mean the collection of all **n-vectors**. This is denoted by $\mathbb{R}^n$.

So, $\boldsymbol{u}$ is an element of $\mathbb{R}^3$, $\boldsymbol{v}$ belongs to $\mathbb{R}^4$ and $\boldsymbol{w} \in \mathbb{R}^5$.

Remark 2.7. The vector $\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$ has four **components**, thus, is a **4-vector** and an element of $\mathbb{R}^4$. Even though the last two components are zero it is not an element of $\mathbb{R}^2$. In fact, no element of $\mathbb{R}^2$ (or $\mathbb{R}^3$) is an element of $\mathbb{R}^4$.

Definition 2.14. Two **n-vectors** $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ are ***equal*** if and only if $a_i = b_i$ for all $i = 1, 2, \ldots, n$, that is, they have identical **components**. In this case we write $\boldsymbol{a} = \boldsymbol{b}$.

The rest of this section will be devoted primarily to the ***algebra*** of $\mathbb{R}^n$ by which we mean that we will define operations called ***(vector) addition*** and ***scalar multiplication*** and we will explicate certain rules, laws or axioms that are satisfied by these operations. We begin with the definition of addition.

Definition 2.15. To ***sum*** or ***add*** two **$n-$vectors** $\boldsymbol{u}, \boldsymbol{v}$ add the corresponding **components**:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}.$$

This is a vector in **n-space, $\mathbb{R}^n$**.

Example 2.2.2.

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

Remark 2.8. As we saw in Section (2.1) if our **vectors** are in **2-space** ($\mathbb{R}^2$) or 3-space ($\mathbb{R}^3$) then **vector addition** satisfies the parallelgram rule: the sum is the vector represented by the diagonal of the parallelgram spanned by the two vectors as shown in the Figure (2.2.1).

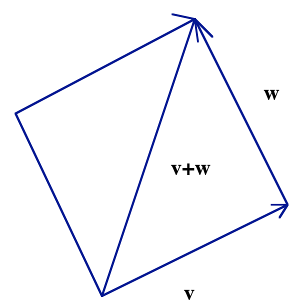


Figure 2.2.1: Parallelogram Rule

Example 2.2.3. When **n-vectors** are **added**, their **components** can have specific meaning. For example, for each state S let $\boldsymbol{p}^S$ be the 111-vector whose i^{th} coordinate is the number of persons in S of age $i = 0, 1, \ldots, 110$, the age population vector for that state. It makes sense to add such vectors - we could add all the vectors in a region, e.g. New England, the South, the Southwest, or even all the states - and get the corresponding age population vector for the region (the nation).

The second operation involves a real number c (a scalar) and an **n-vector** $\boldsymbol{u}$.

Definition 2.16. Let $\boldsymbol{u}$ be an **n-vector** and c a scalar. The ***scalar product or multiple*** $c\boldsymbol{u}$ is obtained by multiplying all the **components** of $\boldsymbol{u}$ by the number (scalar) c. This is a vector in $\mathbb{R}^n$.

$$c\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{pmatrix}.$$

Example 2.2.4.

$$(-3)\begin{pmatrix} -2 \\ 0 \\ 1 \\ -2 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ -3 \\ 6 \\ 3 \\ -9 \end{pmatrix}.$$

Example 2.2.5. Recall, in part two of **Example** (2.2.1) the nutritional vectors $\boldsymbol{v}_{PC} = \begin{pmatrix} 2.5 \\ 18.5 \\ 3 \\ 110 \end{pmatrix}$ recorded the grams of protein, carbohydrates, fats and the calories of a one ounce serving of Maple Pecan clusters and $\boldsymbol{v}_{OF} = \begin{pmatrix} 3 \\ 23 \\ 1 \\ 110 \end{pmatrix}$ provided the same information for a one ounce serving of Oak Flakes cereal. A three-ounce serving of cereal consisting of 1 ounce of Maple Pecan clusters and 2 ounces of Oak Flakes has nutritional vector

$$1\begin{pmatrix} 2.5 \\ 18.5 \\ 3 \\ 110 \end{pmatrix} + 2\begin{pmatrix} 3 \\ 23 \\ 1 \\ 110 \end{pmatrix} = \begin{pmatrix} 8.5 \\ 64.5 \\ 5 \\ 330 \end{pmatrix}$$

Of special importance will be the vector $(-1)\boldsymbol{u}$:

Definition 2.17. The **scalar product** of the **n-vector** $\boldsymbol{u}$ by (-1), $(-1)\boldsymbol{u}$ which we denote by $-\boldsymbol{u}$, is called the ***negative of*** $\boldsymbol{u}$.

Example 2.2.6. Suppose we look at the age population vector of California at different times, for example, on the first of January every year. For example, $\boldsymbol{p}^{Ca}_{2003}$ is the age population vector for California on January 1, 2003 and $\boldsymbol{p}^{Ca}_{2004}$ is the age population vector of California on January 1, 2004. The difference $\boldsymbol{p}^{Ca}_{2004} - \boldsymbol{p}^{Ca}_{2003}$ is the change in the age groups in California from 2003 to 2004. These components can be positive, negative or zero, depending on whether the population of an age group increased, decreased or remained the same.

Also fundamental to our theory is the vector whose components are all zero:

Definition 2.18. The **zero vector** in **n-space, $\mathbb{R}^n$** is the n-vector all of whose components are zero. We will denote this by $\mathbf{0}_n$ or just $\mathbf{0}$ when the n is clear from the context.

Example 2.2.7.

$$\mathbf{0}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \mathbb{R}^2, \mathbf{0}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \in \mathbb{R}^3, \mathbf{0}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \in \mathbb{R}^4.$$

The following sequence of **n-vectors** will play a leading role later in the book. It will then become clear why we refer to it as a basis (to be defined later) and why, among all bases of $\mathbb{R}^n$ it is the **standard** one.

Definition 2.19. ***The Standard Basis of*** $\mathbb{R}^n$

For a given n we will denote by $\boldsymbol{e}^n_i$ the **$n-$vector** which has only one non-zero component, a one, which occurs in the i^{th} row . This is the i^{th} ***standard basis vector of*** $\mathbb{R}^n$***.*** The sequence, $(\boldsymbol{e}^n_1, \boldsymbol{e}^n_2, \ldots, \boldsymbol{e}^n_n)$ is the ***standard basis of*** $\mathbb{R}^n$. When the n is understood from the context we will usually not use the superscript.

Example 2.2.8.

The **standard basis vectors** for $\mathbb{R}^3$ are $\boldsymbol{e}_1^3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{e}_2^3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{e}_3^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

When we fix n and consider the collection of $n-$vectors, $\mathbb{R}^n$, then the following properties hold. These are precisely the conditions for $\mathbb{R}^n$ to be an abstract vector space, a concept which will occupy us in chapter five.

Theorem 2.2.1. **Properties of vector addition and scalar multiplication**

*Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be **n-vectors** and a, b be scalars (real numbers). Then the following properties hold.*

1. $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$. *Associative law*
2. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$. *Commutative law*
3. $\boldsymbol{u} + \boldsymbol{0} = \boldsymbol{u}$. *The **zero vector** is an additive identity*
4. $\boldsymbol{u} + (-\boldsymbol{u}) = \boldsymbol{0}$. *Existence of additive inverses*
5. $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$. *Distributive law*
6. $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$. *Distributive law*
7. $(ab)\boldsymbol{u} = a(b\boldsymbol{u})$.
8. $1\boldsymbol{u} = \boldsymbol{u}$.
9. $0\boldsymbol{u} = \boldsymbol{0}$.

Proof. We prove some of these in the case where $n = 3$. In each instance the general case for an arbitrary n is proved in exactly the same way.

1. Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be **3-vectors** and assume that

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

Then

$$(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \begin{pmatrix} (u_1 + v_1) + w_1 \\ (u_2 + v_2) + w_2 \\ (u_3 + v_3) + w_3 \end{pmatrix}$$

and

$$\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = \begin{pmatrix} u_1 + (v_1 + w_1) \\ u_2 + (v_2 + w_2) \\ u_3 + (v_3 + w_3) \end{pmatrix}.$$

However, since for the addition of real numbers $(u_i + v_i) + w_i = u_i + (v_i + w_i)$ for $i = 1, 2, 3$ it follows that these **3-vectors** are **equal**.

In a similar fashion 2. holds since it will reduce to showing that the **components** of $\boldsymbol{u}+\boldsymbol{v}$ and $\boldsymbol{v}+\boldsymbol{u}$ are **equal**. For example, the first component of $\boldsymbol{u}+\boldsymbol{v}$ and $\boldsymbol{v}+\boldsymbol{u}$ are, respectively u_1+v_1, v_1+u_1 which are the equal.

3. This holds since we are adding 0 to each **component** of $\boldsymbol{u}$ and so leaving $\boldsymbol{u}$ unchanged.

4. The **components** of $\boldsymbol{u}+(-\boldsymbol{u})$ are $u_1-u_1, u_2-u_2, u_3-u_3$ and are therefore all zero.

5. and 6. hold because the distributive law applies to real numbers:

The **components** of $a(\boldsymbol{u}+\boldsymbol{v}), a\boldsymbol{u}+a\boldsymbol{v}$ are, respectively, $a(u_1+v_1), a(u_2+v_2), a(u_2+v_3)$ and $au_1+av_1, au_2+av_2, au_3+av_3$ and these are all equal.

6. This is left as an exercise.

7. The **components** of $(ab)\boldsymbol{u}$ are $(ab)u_1, (ab)u_2, (ab)u_3$. The components of $a(b\boldsymbol{u})$ are $a(bu_1), a(bu_2), a(bu_3)$. $(ab)u_1 = a(bu_1), (ab)u_2 = a(bu_2), (ab)u_3 = a(bu_3)$ since multiplication of real numbers is associative.

8. Here, each **component** is being multiplied by 1 and so is unchanged and therefore $\boldsymbol{u}$ is unchanged.

9. Each **component** of $\boldsymbol{u}$ is being multiplied by 0 and so is 0. Therefore the result is the **zero vector**. $\square$

We now come to one of the most important definitions in the course. Given some **n-vectors**, say $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4$ we can **multiply** these by any scalars we want, say -1, 2, -3, 5 to get the $n-$vectors $-\boldsymbol{u}_1, 2\boldsymbol{u}_2, -3\boldsymbol{u}_3, 5\boldsymbol{u}_4$

Then we can **add** these (and we don't need to worry about parentheses because of the associative law for addition) in order to get $-\boldsymbol{u}_1+2\boldsymbol{u}_2-3\boldsymbol{u}_3+5\boldsymbol{u}_4$.

This is an example of the following concept:

Definition 2.20. Let $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m)$ be a sequence of **$n-$vectors** and $c_1, c_2, \ldots, c_m$ be scalars (real numbers). An $n-$vector of the form

$$c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \ldots c_m\boldsymbol{u}_m$$

is a ***linear combination*** of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m)$. The scalars, $c_1, c_2, \ldots, c_m$ are called the ***coefficients*** of the linear combination.

Example 2.2.9. Suppose a manufacturer has a choice of using four chemicals to produce fertilizer, which has precursors nitrogen, phosphoric acid and potash. Suppose the composition of these four chemical as a percentage of these ingredients (by weight) are given by the following **3-vectors**:

$$\boldsymbol{c}_1 = \begin{pmatrix} 0.20 \\ 0.12 \\ 0 \end{pmatrix}, \boldsymbol{c}_2 = \begin{pmatrix} 0.25 \\ 0.05 \\ .05 \end{pmatrix}, \boldsymbol{c}_3 = \begin{pmatrix} 0 \\ 0.06 \\ 0.15 \end{pmatrix}, \boldsymbol{c}_4 = \begin{pmatrix} 0.30 \\ 0.07 \\ 0.10 \end{pmatrix}$$

If we use x_i pounds of chemical $i = 1, 2, 3, 4$ then the amount of each ingredient in the mixture is given by the **components** of the **linear combination**

$$x_1\boldsymbol{c}_1 + x_2\boldsymbol{c}_2 + x_3\boldsymbol{c}_3 + x_4\boldsymbol{c}_4$$

Notation

1. If $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k)$ is a sequence of **n-vectors** then $(\boldsymbol{u}_1\ \boldsymbol{u}_2\ \ldots\ \boldsymbol{u}_k)$ will denote the $n \times k$ matrix whose columns are the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$.

2. Let A be an $n \times k$ matrix and $\boldsymbol{b}$ an **n-vector**. By $[A|\boldsymbol{b}]$ we shall mean the matrix A augmented by the column $\boldsymbol{b}$.

As we will see in the example below when given a sequence $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k)$ of **n-vectors** and an n-vector $\boldsymbol{u}$ the following two questions are logically equivalent:

1. Can the **n-vector** $\boldsymbol{u}$ be expressed as a **linear combination** of $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$?

2. Is the **linear system** with **augmented matrix** $[(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k)|\boldsymbol{u}]$ **consistent**?

We previously demonstrated how to answer second question: Use **Gaussian elimination** to obtain a **row echelon form** of the augmented matrix $[(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k)|\boldsymbol{u}]$. The linear system is **consistent** if and only if the last column is not a **pivot column**.

When the linear system is **consistent** and we want to express $\boldsymbol{u}$ as a **linear combination** of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k)$ then we can use the **reduced row echelon form** to find one. In turn, we can interpret a **linear system** as a vector equation.

Example 2.2.10. Determine whether the **3-vector** $\boldsymbol{u} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ is a **linear combination** of the vectors $\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$.

The vector $\boldsymbol{u}$ is a **linear combination** if and only if there are scalars c_1, c_2, c_3 so that $c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + c_3\boldsymbol{u}_3 = \boldsymbol{u}$, that is,

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

This becomes

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

This is the same as the **linear system**

$$\begin{array}{rcrcrcr} c_1 & - & c_2 & + & c_3 & = & 1 \\ c_1 & + & 3c_2 & & & = & -2 \\ -2c_1 & - & 2c_2 & - & c_3 & = & 1 \end{array}$$

The **augmented matrix of this system** is

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

Notice that the columns of this matrix are the vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ and $\boldsymbol{u}$.
Applying **Gaussian elimination** we obtain an **echelon form**:

$$\to \begin{pmatrix} 1 & -1 & 1 & | & 1 \\ 0 & 4 & -1 & | & -3 \\ 0 & -4 & 1 & | & 3 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 1 & | & 1 \\ 0 & 4 & -1 & | & -3 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

At this point we can say that the linear system is **consistent** and so the answer is yes, $\boldsymbol{u}$ is a **linear combination** of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$. We proceed to actually find such linear combinations by continuing the use of **Gaussian elimination** to find the **reduced row echelon form** of the **augmented matrix** of the **linear system** :

$$\to \begin{pmatrix} 1 & -1 & 1 & | & 1 \\ 0 & 1 & -\frac{1}{4} & | & -\frac{3}{4} \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & \frac{3}{4} & | & \frac{1}{4} \\ 0 & 1 & -\frac{1}{4} & | & -\frac{3}{4} \\ 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

The final matrix is the **augmented matrix** of the following linear system (which is **equivalent** to the original one):

$$\begin{array}{rcrcrcr} c_1 & & & + & \frac{3}{4}c_3 & = & \frac{1}{4} \\ & & c_2 & - & \frac{1}{4} & = & -\frac{3}{4} \end{array}$$

This system has two **leading variables and one free variable** and consequently we can express $\boldsymbol{u}$ as a linear combination in infinitely many ways. The parametric version (general solution) is

$$\begin{aligned} c_1 &= \frac{1}{4} - \frac{3}{4}t \\ c_2 &= -\frac{3}{4} + \frac{1}{4}t \\ c_3 &= t. \end{aligned}$$

As a particular solution we can set $t = 0$ to get $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ -\frac{3}{4} \\ 0 \end{pmatrix}$.

We check that this is, indeed, a solution to the system:

$$\frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} - \frac{3}{4}\begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} - (-\frac{3}{4}) \\ \frac{1}{4} - \frac{9}{4} \\ -\frac{2}{4} + \frac{6}{4} \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \boldsymbol{u}.$$

In the course of the above example we saw that the question of whether an **n-vector** $\boldsymbol{u}$ is a **linear combination** of the sequence of $n-$vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is equivalent to whether or not the **linear system** with **augmented matrix** $((\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_k) \mid \boldsymbol{u})$ is **consistent** and each **solution** $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ to the **linear system** gives a **linear combination** $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_k\boldsymbol{v}_k = \boldsymbol{u}$.

On the other hand the equivalence goes the other way as illustrated in the following:

Example 2.2.11. Consider the following **linear system**

$$\begin{array}{rcrcrcrcr} -2x & + & 3y & & & + & 4w & = & -6 \\ 3x & - & y & + & 4z & + & 3w & = & 4 \\ x & & & - & 8z & - & 10w & = & 11 \\ -5x & + & 2y & + & 3z & & & = & 0 \end{array}.$$

Because two **n-vectors** are equal if and only if corresponding **components** are equal this can be written as an equation between vectors:

$$\begin{pmatrix} -2x & + & 3y & & & + & 4w \\ 3x & - & y & + & 4z & + & 3w \\ x & & & - & 8z & - & 10w \\ -5x & + & 2y & + & 3z & & \end{pmatrix} = \begin{pmatrix} -6 \\ 4 \\ 11 \\ 0 \end{pmatrix}.$$

On the other hand, by the definition of **vector addition**, the left hand side can be expressed as

$$\begin{pmatrix}-2x\\3x\\x\\-5x\end{pmatrix}+\begin{pmatrix}3y\\-y\\0\\2y\end{pmatrix}+\begin{pmatrix}0\\4z\\-8z\\3z\end{pmatrix}+\begin{pmatrix}4w\\3w\\-10w\\0\end{pmatrix}$$

And then by the definition of **scalar multiplication** this becomes

$$x\begin{pmatrix}-2\\3\\1\\-5\end{pmatrix}+y\begin{pmatrix}3\\-1\\0\\2\end{pmatrix}+z\begin{pmatrix}0\\4\\-8\\3\end{pmatrix}+w\begin{pmatrix}4\\3\\-10\\0\end{pmatrix}=\begin{pmatrix}-6\\4\\11\\0\end{pmatrix}.$$

And so the **consistency** of the original **linear system** is equivalent to the question of whether the columns of constants $\begin{pmatrix}-6\\4\\11\\0\end{pmatrix}$ is a **linear combination** of the columns of the **coefficient matrix of the system**.

This correspondence between the **consistency** and **solutions** of a **linear system** and the question of whether an **n-vector** is a **linear combination** of a sequence of **n-vectors** is so important that we state it formerly as a theorem:

Theorem 2.2.2. *Let A be an $n \times k$ **matrix** with **columns** the **n-vectors** $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ and assume that $\boldsymbol{b}$ is an **n-vector**. Then the **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$ is **consistent** if and only if the vector $\boldsymbol{b}$ is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$. Moreover, $\begin{pmatrix}c_1\\c_2\\\vdots\\c_k\end{pmatrix}$ is a **solution** to the linear system if and only if $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \boldsymbol{b}$.*

What You Can Now Do

1. Given a sequence of **n-vectors**, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$, and an $n-$vector $\boldsymbol{u}$, determine if $\boldsymbol{u}$ is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.

2. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of **n-vectors** and an n-vector $\boldsymbol{u}$ which is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ find scalars $c_1, \dots, c_k$ such that $c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \boldsymbol{u}$.

Method (How do do it)

Method 2.2.1. Determine if an **n-vector** $\boldsymbol{u}$ is a **linear combination** of the sequence of $n-$vectors $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k)$.

Let A be the $n \times k$ matrix with columns the n-vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$. Determine if the **linear system** with **augmented matrix** $[A|\boldsymbol{u}]$ is **consistent or inconsistent**. If it is consistent then the vector $\boldsymbol{u}$ is a **linear combination** of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k)$. Otherwise it is not.

Example 2.2.12.

Determine if the **4-vector** $\begin{pmatrix} 1 \\ 1 \\ -2 \\ 1 \end{pmatrix}$ is a **linear combination** of the following sequence of vectors.

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 2 \\ -3 \\ 0 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 5 \\ -4 \\ -3 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} -1 \\ -2 \\ 5 \\ -2 \end{pmatrix}, \boldsymbol{u}_4 = \begin{pmatrix} 3 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

We form the **augmented matrix**

$$\begin{pmatrix} 1 & 2 & -1 & 3 & | & 1 \\ 2 & 5 & -2 & -1 & | & 1 \\ -3 & -4 & 5 & -1 & | & -2 \\ 0 & -3 & -2 & -1 & | & 1 \end{pmatrix}$$

and determine if it is the matrix of a **consistent linear system** by applying **Gaussian elimination** to obtain an **echelon form**.

$$\to \begin{pmatrix} 1 & 2 & -1 & 3 & | & 1 \\ 0 & 1 & 0 & -7 & | & -1 \\ 0 & 2 & 2 & 8 & | & 1 \\ 0 & -3 & -2 & -1 & | & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -1 & 3 & | & 1 \\ 0 & 1 & 0 & -7 & | & -1 \\ 0 & 0 & 2 & 22 & | & 3 \\ 0 & 0 & -2 & -22 & | & -2 \end{pmatrix} \to$$

$$\left(\begin{array}{cccc|c} 1 & 2 & -1 & 3 & 1 \\ 0 & 1 & 0 & -7 & -1 \\ 0 & 0 & 2 & 22 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

Since the last column is a **pivot column** , we see the linear system is **inconsistent** and therefore the **4-vector** $\begin{pmatrix} 1 \\ 1 \\ -2 \\ 1 \end{pmatrix}$ is not a **linear combination** of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4)$.

Example 2.2.13. Determine if the **4-vector** $\boldsymbol{u} = \begin{pmatrix} 3 \\ -4 \\ 3 \\ -2 \end{pmatrix}$ is a **linear combination** of the sequence $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ where

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 3 \\ -2 \\ -2 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 5 \\ -4 \\ -3 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} -1 \\ -3 \\ -1 \\ 5 \end{pmatrix}$$

We form the matrix $(\boldsymbol{u}_1\ \boldsymbol{u}_2\ \boldsymbol{u}_3 \mid \boldsymbol{u})$ which is equal to

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

and determine if the **linear system** with this as its **augmented matrix** is **consistent**. In order to do so, we apply **Gaussian elimination** to obtain an **echelon form** and then determine if the last column is a **pivot column**. We proceed to apply the Gaussian elimination algorithm:

$$\to \left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & -1 & 0 & -13 \\ 0 & 0 & -3 & 9 \\ 0 & 1 & 3 & 4 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & -1 & 0 & -13 \\ 0 & 0 & -3 & 9 \\ 0 & 0 & 3 & -9 \end{array}\right) \to$$
$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & -1 & 0 & -13 \\ 0 & 0 & -3 & 9 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Thus, the linear system is consistent and the vector $\begin{pmatrix} 3 \\ -4 \\ 3 \\ -2 \end{pmatrix}$ is a linear combination of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$.

Method 2.2.2. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of n-vectors and an n-vector $\boldsymbol{u}$ which is a linear combination of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ find scalars $c_1, \dots, c_k$ such that $c_1\boldsymbol{v}_1 + \dots + c_k\boldsymbol{v}_k = \boldsymbol{u}$.

As in Method (2.2.1) form the $n \times k-$matrix A with columns the n-vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k$ and apply Gaussian elimination to the augmented matrix $[A|\boldsymbol{u}]$ to obtain its reduced row echelon form. Since $\boldsymbol{u}$ is a linear combination of the sequence $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k)$ the linear system with this augmented matrix is consistent and any solution will give an expression of $\boldsymbol{u}$ as linear combination of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k)$.

Example 2.2.14. We continue with Example (2.2.13). There we determined that $\boldsymbol{u} = \begin{pmatrix} 3 \\ -4 \\ 3 \\ -2 \end{pmatrix}$ is a linear combination of the sequence of vectors $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ where

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 3 \\ -2 \\ -2 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 5 \\ -4 \\ -3 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} -1 \\ -3 \\ -1 \\ 5 \end{pmatrix}$$

We use Gaussian elimination to obtain the reduced echelon form of the matrix $((\boldsymbol{u}_1\ \boldsymbol{u}_2\ \boldsymbol{u}_3) \mid \boldsymbol{u})$. We then find solutions to the linear system with augmented matrix $((\boldsymbol{u}_1\ \boldsymbol{u}_2\ \boldsymbol{u}_3) \mid \boldsymbol{u})$ and then express $\boldsymbol{u}$ as a linear combination of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$. We proceed to apply Gaussian elimination, starting with the echelon form obtained in Example (2.2.13):

$$\begin{pmatrix} 1 & 2 & -1 & | & 3 \\ 0 & -1 & 0 & | & -13 \\ 0 & 0 & -3 & | & 9 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -1 & | & 3 \\ 0 & -1 & 0 & | & -13 \\ 0 & 0 & 1 & | & -3 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & | & 0 \\ 0 & -1 & 0 & | & -13 \\ 0 & 0 & 1 & | & -3 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 0 & | & 0 \\ 0 & 1 & 0 & | & 13 \\ 0 & 0 & 1 & | & -3 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & -26 \\ 0 & 1 & 0 & | & 13 \\ 0 & 0 & 1 & | & -3 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

We see in this case that there is a unique solution, $\begin{pmatrix} -26 \\ 13 \\ -3 \end{pmatrix}$. We check:

$$(-26)\boldsymbol{u}_1 + 13\boldsymbol{u}_2 + (-3)\boldsymbol{u}_3 =$$

$$(-26)\begin{pmatrix} 1 \\ 3 \\ -2 \\ -2 \end{pmatrix} + 13\begin{pmatrix} 2 \\ 5 \\ -4 \\ -3 \end{pmatrix} + (-3)\begin{pmatrix} -1 \\ -3 \\ -1 \\ 5 \end{pmatrix} =$$

$$\begin{pmatrix} -26 \\ -78 \\ 52 \\ 52 \end{pmatrix} + \begin{pmatrix} 26 \\ 65 \\ -52 \\ -39 \end{pmatrix} + \begin{pmatrix} 3 \\ 9 \\ 3 \\ -15 \end{pmatrix} =$$

$$\begin{pmatrix} -26+26+3 \\ -78+65+9 \\ 52-52+3 \\ 52-39-15 \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \\ 3 \\ -2 \end{pmatrix}$$

Exercises

In exercises 1-10 let $\boldsymbol{x} = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$, $\boldsymbol{y} = \begin{pmatrix} 0 \\ -3 \\ 2 \end{pmatrix}$, $\boldsymbol{z} = \begin{pmatrix} -4 \\ 3 \\ 0 \end{pmatrix}$, $\boldsymbol{u} = \begin{pmatrix} 1 \\ -2 \\ 5 \\ 0 \end{pmatrix}$,

$\boldsymbol{v} = \begin{pmatrix} 2 \\ -3 \\ 4 \\ 1 \end{pmatrix}$, $\boldsymbol{w} = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{3}{4} \\ -1 \end{pmatrix}$. Compute the following:

1. a) $\boldsymbol{x} + \boldsymbol{y}$
b) $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$
c) $\boldsymbol{y} + \boldsymbol{z}$
d) $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$

2. a) $\boldsymbol{x} + \boldsymbol{z}$
b) $\boldsymbol{z} + \boldsymbol{x}$

3. a) $(-2)\boldsymbol{x}$
b) $3\boldsymbol{y}$

c) $2\boldsymbol{x} - \boldsymbol{y} - 4\boldsymbol{z}$

4. a) $\boldsymbol{x} + \boldsymbol{u}$
b) $(\boldsymbol{x} + \boldsymbol{u}) + \boldsymbol{z}$
c) $\boldsymbol{u} + \boldsymbol{z}$
d) $\boldsymbol{x} + (\boldsymbol{u} + \boldsymbol{z})$

5. a) $\boldsymbol{u} + \boldsymbol{v}$
b) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$
c) $\boldsymbol{v} + \boldsymbol{w}$
d) $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$

6. a) $\boldsymbol{u} + \boldsymbol{w}$
b) $\boldsymbol{w} + \boldsymbol{u}$

7. a) $(-5)\boldsymbol{u}$
b) $3\boldsymbol{v}$
c) $2\boldsymbol{u} - \boldsymbol{v} + 2\boldsymbol{w}$

8. a) $5\boldsymbol{x}$
b) $5\boldsymbol{y}$
c) $5\boldsymbol{x} + 5\boldsymbol{y}$
d) $5(\boldsymbol{x} + \boldsymbol{y})$

9. a) $(-3)\boldsymbol{v}$
b) $7\boldsymbol{v}$
c) $(-3)\boldsymbol{v} + 7\boldsymbol{v}$
d) $(-3 + 7)\boldsymbol{v} = 4\boldsymbol{v}$

10. a) $2\boldsymbol{z}$
b) $(-5)(2\boldsymbol{z})$
c) $(-10)\boldsymbol{z}$

In exercises 11-14 let $\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 3 \\ 8 \\ 7 \end{pmatrix}$. Determine if the given **3-vector** $\boldsymbol{u}$ is a **linear combination** of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. If so, give an example of scalars a_1 and a_2 such that $\boldsymbol{u} = a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2$. See **Method** (2.2.1) and **Method** (2.2.2).

11. $\boldsymbol{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

12. $\boldsymbol{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

13. $\boldsymbol{u} = \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix}$

14. $\boldsymbol{u} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

In exercises 15-18 let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \\ 4 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 3 \\ 5 \\ -2 \\ 10 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 4 \\ -4 \\ 9 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ -1 \\ -2 \\ 2 \end{pmatrix}$.

Determine if the given **4-vector** $\boldsymbol{v}$ is a **linear combination** of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$. If so, give an example of scalars c_1, c_2, c_3, c_4 such that $\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4$. See **Method** (2.2.1) and **Method** (2.2.2).

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

16. $\begin{pmatrix} 4 \\ 7 \\ -2 \\ 9 \end{pmatrix}$

17. $\begin{pmatrix} 0 \\ 3 \\ 1 \\ 2 \end{pmatrix}$

18. $\begin{pmatrix} 2 \\ 1 \\ -3 \\ 2 \end{pmatrix}$

In exercises 19 - 24 answer true or false and give an explanation.

19. The vector $\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ belongs to $\mathbb{R}^2$.

20. Every vector in $\mathbb{R}^2$ is also in $\mathbb{R}^3$.

21. The **sum** of two **3-vectors** is a 6-vector.

22. The vector $2\boldsymbol{v}_1 - 3\boldsymbol{v}_2$ is a **linear combination** of the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ from $\mathbb{R}^n$.

23. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are **n-vectors** and $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{u}$ then $\boldsymbol{v} = \mathbf{0}_n$, the **zero vector** in $\mathbb{R}^n$.

24. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are **n-vectors** and $\boldsymbol{u} + \boldsymbol{v} = \mathbf{0}_n$ then $\boldsymbol{v} = -\boldsymbol{u}$, the **negative** of $\boldsymbol{u}$.

Challenge Exercises (Problems)

1. Clearly every **3-vector** in $\mathbb{R}^3$ is a **linear combination** of the **standard basis** of $\mathbb{R}^3$:

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

a) Verify that each of the **standard basis vectors**, $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ is a **linear combination** of the vectors

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 5 \\ 7 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 3 \\ 6 \\ 8 \end{pmatrix}.$$

b) Explain why every **3-vector** must be a **linear combination** of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$.

2. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 5 \\ -2 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} -2 \\ -5 \\ -4 \\ 3 \end{pmatrix}$

a) Let $\boldsymbol{v} = \begin{pmatrix} 2 \\ 2 \\ 7 \\ -1 \end{pmatrix}$. Verify that $\boldsymbol{v} = 2\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3$.

b) Explain why the **linear system** with **augmented matrix**

$$\left(\begin{array}{ccc|c} 1 & 2 & -2 & 2 \\ 2 & 3 & -5 & 2 \\ 3 & 5 & -4 & 7 \\ -1 & -2 & 3 & -1 \end{array}\right)$$

is **consistent**.

c) Explain why there must exist a **4-vector** $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}$ such that the **linear system** with **augmented matrix**

$$\begin{pmatrix} 1 & 2 & -2 & | & b_1 \\ 2 & 3 & -5 & | & b_2 \\ 3 & 5 & -4 & | & b_3 \\ -1 & -2 & 3 & | & b_4 \end{pmatrix}$$

is **inconsistent**.

d) Explain why there must exist a **4-vector** $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}$ such that $\boldsymbol{b}$ is not a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

3. Let

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}.$$

a) Verify that each of $\boldsymbol{v}_i, i = 1, 2, 3, 4$ belongs to

$$S = \{\boldsymbol{v} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \in \mathbb{R}^4 | a + b + c + d = 0\}.$$

b) Explain why there must exist a **4-vector** $\boldsymbol{b}$ which is not a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$. (Hint: Demonstrate that every linear combination of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_4$ lies in S.) In particular, explain why $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ is not a linear combination of these vectors.

Quiz Solutions

1.

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

Not right, see **Method** (1.2.3)

2. The **pivot columns** are 1,3 and 4. The following is an **echelon form** of the matrix:

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

Not right, see the definition of a **pivot columns of a matrix**

3. The number of **solutions** of this system is c) infinite
The general solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 + 2s + t \\ s \\ -2 - 2t \\ t \end{pmatrix}$$

and you can choose s, t arbitrarily. Alternatively, the system is **consistent** and not every column (other than the augmented column) is a **pivot column** so that there are **free variables**. Not right, see **Theorem** (1.2.3)

4. a) b) c) and d) give solutions. e) is not a solution.

Not right, see definition of **solution of a linear system**

2.3. The Span of a Sequence of Vectors

Am I Ready for This Material

In order to follow this section and master its ideas and methods you need to be familiar with many previously defined concepts. **Those that are used extensively in this section are**:

linear system

solution of a linear system

solution set of a linear system

consistent linear system

inconsistent linear system

equivalent linear systems

echelon form of a linear system

leading variable

free variable

matrix

augmented matrix of a linear system

row equivalence of matrices

matrix in row echelon form

matrix in reduced row echelon form

echelon form of a matrix

reduced echelon from of a matrix

pivot positions of a matrix

pivot columns of a matrix

an $n-$vector

equality of $n-$vectors

$n-$space, $\mathbb{R}^n$

adddition of $n-$vectors

scalar multiplication of an $n-$vector

negative of a vector

the zero vector

the standard basis for $\mathbb{R}^n$

linear combination of vectors

You will also need to be able to execute several procedures previously described. **The following are algorithms that we will use in this section**:

Gaussian elimination

procedure for solving a linear system

Quiz

Let $\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 3 \\ 5 \\ 7 \\ 2 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \end{pmatrix}$; and $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 7 \\ -6 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 8 \\ -3 \end{pmatrix}$.

1. Determine if the **4-vector** $\begin{pmatrix} 2 \\ 4 \\ 7 \\ 2 \end{pmatrix}$ is a **linear combination** of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ and if so, find such a linear combination.

2. Express each of the **standard basis vectors** of $\mathbb{R}^3$ as a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

3. Let $\boldsymbol{u} = \begin{pmatrix} -3 \\ 2 \\ 5 \end{pmatrix}$. Express $\boldsymbol{u}$ as a **linear combination** of the **standard basis vectors** $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ of $\mathbb{R}^3$ and as a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$.

Quiz Solutions

New Concepts

Several fundamental concepts are introduced in this section. It is extremely important to genuinely understand and master them. These are:

span of a sequence of **vectors** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ from $\mathbb{R}^n$

spanning sequence of $\mathbb{R}^n$

subspace of $\mathbb{R}^n$

spanning sequence of a subspace V of $\mathbb{R}^n$

In addition we will define

column space of a matrix

row space of a matrix

Theory (Why it Works)

Warning: The material introduced here is fundamental and will be used in nearly every subsequent section of the book

We previously introduced the concept of a **linear combination** of a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of **n-vectors**. We will be especially interested in describing the set of all linear combinations of these vectors and finding conditions for when it is all of $\mathbb{R}^n$. Because the concept of the set of all linear combinations of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is so important and we refer to it so often we assign a special name to it:

Definition 2.21. Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be a sequence of $n-$**vectors**. The set (collection) of all **linear combinations** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is called the **span** of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ and is denoted by

$$Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k).$$

If $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ are vectors from $\mathbb{R}^n$ and $Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \mathbb{R}^n$ then we say that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is a ***spanning sequence*** of $\mathbb{R}^n$ or simply that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ ***spans*** $\mathbb{R}^n$.

In what follows we will draw many consequences from these definitions.

Before proceeding to the theorems we first do a couple of examples.

Example 2.3.1. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$. Determine whether $\boldsymbol{u} = \begin{pmatrix} 2 \\ -1 \\ 2 \\ -3 \end{pmatrix}$ is in the span of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$?

This is a question we encountered before when we introduced the notion of **linear**

combination. We are just phrasing the question differently, but ultimately we are asking: is the **4-vector** $\boldsymbol{u}$ a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$?

This is the same as asking if there are scalars c_1, c_2, c_3 such that

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \boldsymbol{u}?$$

We have previously seen that this reduces to determining if the **linear system** with **augmented matrix** $(\boldsymbol{v}_1\boldsymbol{v}_2\boldsymbol{v}_3|\boldsymbol{u})$ is **consistent**.

The **matrix** of the linear system is $\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 1 & 1 & -1 & -1 \\ 1 & -2 & 0 & 2 \\ -3 & 0 & 0 & -3 \end{array}\right)$.

We solve by using **Gaussian elimination** to obtain the **reduced echelon form** of the matrix.

The matrix we obtain is $\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{array}\right)$. The last column is not a **pivot column** so the system is **consistent** by **Theorem** (1.2.1) and, consequently, $\boldsymbol{u}$ is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Moreover, we can see that there is a unique **solution**, namely, $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -\frac{1}{2} \\ \frac{3}{2} \end{pmatrix}$. We check and see that it works:

$$(1)\begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix} + (-\frac{1}{2})\begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \end{pmatrix} + (\frac{3}{2})\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 2 \\ -3 \end{pmatrix}$$

as required.

Example 2.3.2. Is it possible that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = \mathbb{R}^4$?

This would require that for every **4-vector** $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix}$ the **linear system** with **augmented matrix** $(\boldsymbol{v}_1\boldsymbol{v}_2\boldsymbol{v}_3|\boldsymbol{b})$ is **consistent**. We shall see this is not possible.

The **coefficient matrix of the linear system** is the following 4×3 matrix.

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

We know that any **echelon form of this matrix** will have a **zero row**: since there are only three columns there are at most three **pivot columns**, allowing for at most three **pivot positions** and hence, at most three non-zero rows. We find an **echelon form** by using **Gaussian elimination**:

Using $R_2 \to (-1)R_1 + R_2, R_3 \to (-1)R_1 + R_3, R_4 \to 3R_1 + R_4$ we obtain the matrix

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

Then if we apply $R_2 \leftrightarrow R_4$ we get the matrix

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

After rescaling the second row by $R_2 \to \frac{1}{3}R_2$ we obtain

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

By applying the elimination operation $R_3 \to 3R_2 + R_3$ we get

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

Rescaling the third row with $R_3 \to \frac{1}{2}R_3$ yields the matrix

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

Finally, by using $R_4 \to 2R_3 + R_4$ we obtain the **echelon form**

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

As argued a priori, there is a **zero row** in this **echelon form**. It follows that the **linear system** with **augmented matrix**

$$\begin{pmatrix} 1 & 1 & 1 & | & 0 \\ 0 & 1 & 1 & | & 0 \\ 0 & 0 & 1 & | & 0 \\ 0 & 0 & 0 & | & 1 \end{pmatrix} \tag{2.9}$$

is **inconsistent**.

If we reverse the **elementary row operations** used to obtain the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ from A and apply them to the matrix in (2.9) we get the augmented matrix

$$\begin{pmatrix} 1 & 1 & 1 & | & 0 \\ 1 & 1 & -1 & | & 1 \\ 1 & -2 & 0 & | & 0 \\ -3 & -2 & 0 & | & 0 \end{pmatrix} \tag{2.10}$$

Since the matrices (2.9) and (2.10) are **row equivalent** the linear systems they represent are **equivalent** as well. Since the system represented by (2.9) is **inconsistent** it follows that the system with augmented matrix (2.10) is also inconsistent. Consequently, the **4-vector** $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ is not in the **span** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Therefore, $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) \neq \mathbb{R}^4$.

From the example we can extract the following general result:

Theorem 2.3.1. **A criteria for spanning**

Let A be an **n x k matrix** *with columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$. Then the following are equivalent:*

1. For every **n-vector** $\boldsymbol{b}$ *the* **system of linear equations** *with* **augmented matrix** $[A|\boldsymbol{b}]$ *is* **consistent***;*

2. Every **n-vector** $\boldsymbol{b}$ *is a* **linear combination** *of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.*

3. A has n **pivots positions** *(equivalently, every row contains a pivot position).*

Proof. Suppose $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are vectors from $\mathbb{R}^n$. Let A be the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_k)$. Let A' be an **echelon form** of A (for example, the **reduced echelon form of A**). Assume A' has a **zero row** (equivalently, fewer than n **pivots positions**). We claim that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k) \neq \mathbb{R}^n$ (the **span** of $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is not all of $\mathbb{R}^n$).

Let $E_1, E_2, \dots, E_m$ be a sequence of **elementary row operations** which takes A to A'. Let $F_1, F_2, \dots, F_m$ be the inverse of these operations. (The elementary operation $R_i \leftrightarrow R_j$ is its own inverse, the inverse of $R_i \to cR_i$ is $R_i \to \frac{1}{c}R_i$ and the inverse to $R_j \to cR_i + R_j$ is $R_j \to (-c)R_i + R_j$.) Then the sequence of operations $F_m, F_{m-1}, \dots, F_1$ takes A' to A.

The matrix A' has a zero row and therefore the **linear system** with **augmented matrix** $[A'|\boldsymbol{e}_n^n]$ is **inconsistent** since the last column is a **pivot column** ($\boldsymbol{e}_n^n$ is the n^{th} vector of the **standard basis**). Applying the sequence $F_m, F_{m-1}, \dots, F_1$ to this augmented matrix we get an **equivalent** augmented matrix $[A|\boldsymbol{b}]$ where $\boldsymbol{b}$ is the vector obtained by applying the sequence $F_m, F_{m-1}, \dots, F_1$ to the vector $\boldsymbol{e}_n^n$. Thus, the **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$ is equivalent to the linear system with augmented matrix $[A'|\boldsymbol{e}_n^n]$. Since the latter is **inconsistent**, so is the former. It follows that the vector $\boldsymbol{b}$ is not in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ by **Theorem** (2.2.2).

On the other hand, suppose A' does not have a **zero row** (the same as every row contains a **pivot position**). Let $\boldsymbol{b} \in \mathbb{R}^n$ and consider the **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$. The matrix $[A|\boldsymbol{b}]$ is **row equivalent** to a matrix $[A'|\boldsymbol{b}']$ (which is in echelon form) obtained by the same sequence of row operations which transform A to A'. Since A' does not have a zero row the linear sytem with augmented matrix $[A'|\boldsymbol{b}']$ is **consistent** (the last column is not a **pivot column**) whence $[A|\boldsymbol{b}]$ is consistent. This implies that $\boldsymbol{b}$ is a **linear combination** of the sequence of columns, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$, of A by **Theorem** (2.2.2). Since $\boldsymbol{b}$ is arbitrary in $\mathbb{R}^n$ this, in turn, implies that $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is a **spanning sequence of $\mathbb{R}^n$**

□

As a consequence of **Theorem** (2.3.1) we have the following result which gives a necessary condition for a sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ of **n-vectors** to **span** in $\mathbb{R}^n$.

Theorem 2.3.2. **A necessary condition for spanning $\mathbb{R}^n$**

*Assume $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is a **spanning sequence** of $\mathbb{R}^n$. Then $k \geq n$.*

Proof. The matrix $A = (\boldsymbol{v}_1 \boldsymbol{v}_2 \dots \boldsymbol{v}_k)$ has a **pivot position** in each row. This implies that there are as least as many columns (k) as there are rows (n). □

We now prove some fundamental theorems about the **span** of a sequence of **n-vectors** . The first theorem simply says if you have a sequence of vectors and you add a vector to it, then any vector which is in the span of the first set is in the span of the expanded set. Before we state the theorem precisely we introduce a simple concept from set theory:

Definition 2.22. Assume X and Y are sets (collections of objects). If every object that is in X also belongs to Y then we says that X is a ***subset*** of Y and write $X \subseteq Y$.

We now proceed to our next result

Lemma 2.3.3. *Let* $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ *be a sequence of* **n-vectors**. *Then* $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{k-1}) \subseteq Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_k)$.

Proof. Any **linear combination** $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_{k-1}\boldsymbol{v}_{k-1}$ of $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_{k-1})$ is also a linear combination of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ since we can just set $c_k = 0$ and we have $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_{k-1}\boldsymbol{v}_{k-1} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_{k-1}\boldsymbol{v}_{k-1} + 0\boldsymbol{v}_k$, which is in $Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$. $\square$

A conseqence of this is that if X, Y are finite sequences of **n-vectors** and every vector in X occurs in Y then $Span\ X \subseteq Span\ Y$.

This next theorem gives some insight to the kinds of subsets of $\mathbb{R}^n$ we obtain when we take the **span** of a sequence of vectors $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$.

Theorem 2.3.4. *Let* $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ *be a sequence of* **n-vectors** *and set* $V = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$. *Then the following hold:*

1. *If* $\boldsymbol{u}, \boldsymbol{w} \in V$ *then* $\boldsymbol{u} + \boldsymbol{w} \in V$.
2. *If* $\boldsymbol{u} \in V, c \in \mathbb{R}$ *(a scalar) then* $c\boldsymbol{u} \in V$.

Before proceeding to the proof we introduce an extremely important definition:

Definition 2.23. A nonempty subset V of **vectors** from $\mathbb{R}^n$ is called a ***subspace*** if the following conditions hold:

1. If $\boldsymbol{u}, \boldsymbol{w} \in V$ then $\boldsymbol{u} + \boldsymbol{w} \in V$.
2. If $\boldsymbol{u} \in V, c \in \mathbb{R}$ (a scalar) then $c\boldsymbol{u} \in V$.

When V is a subset of $\mathbb{R}^n$ and the first condition of **Definition** (2.23) holds we say that V is "closed under addition." When the second condition holds we say that V is "closed under scalar multiplication."

With this definition **Theorem** (2.3.4) can be restated:

Theorem 2.3.5. *If* $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ *is sequence of* **n-vectors** *then* $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ *is a subspace of* $\mathbb{R}^n$.

Before going directly to the proof let us see what the main idea is. For the purpose of exposition let us suppose that $k = 3$. If we have two **n-vectors** $\boldsymbol{u}, \boldsymbol{w}$ which are in the **span** this just means they are **linear combinations** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Examples are $\boldsymbol{u} = -2\boldsymbol{v}_1 + 5\boldsymbol{v}_2 + 3\boldsymbol{v}_3, \boldsymbol{w} = -7\boldsymbol{v}_1 - 3\boldsymbol{v}_2 + 6\boldsymbol{v}_3$. Adding these two vectors we obtain

$$\begin{aligned} \boldsymbol{u} + \boldsymbol{v} &= (-2\boldsymbol{v}_1 + 5\boldsymbol{v}_2 + 3\boldsymbol{v}_3) + (7\boldsymbol{v}_1 - 3\boldsymbol{v}_2 + 6\boldsymbol{v}_3) \\ &= [(-2) + 7]\boldsymbol{v}_1 + [5 + (-3)]\boldsymbol{v}_2 + [3 + 6]\boldsymbol{v}_3 \\ &= 5\boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 9\boldsymbol{v}_3 \end{aligned}$$

which is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ and therefore in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. We now prove the theorem in general.

Proof. We can write $\boldsymbol{u} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k$ and $\boldsymbol{w} = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_k\boldsymbol{v}_k$ for some scalars $a_1, a_2, \ldots, a_k$ and $b_1, b_2, \ldots, b_k$. Now

$$\begin{aligned} \boldsymbol{u} + \boldsymbol{w} &= (a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k) + (b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_k\boldsymbol{v}_k) \\ &= a_1\boldsymbol{v}_1 + b_1\boldsymbol{v}_1 + a_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k + b_k\boldsymbol{v}_k \\ &= (a_1 + b_1)\boldsymbol{v}_1 + (a_2 + b_2)\boldsymbol{v}_2 + \cdots + (a_k + b_k)\boldsymbol{v}_k. \end{aligned}$$

The proof of the second part is left as a **challenge exercise**. □

At this point we know that the **span of a sequence of vectors** in $\mathbb{R}^n$ is a **subspace** but what about the converse of this statement: If V is a subspace of $\mathbb{R}^n$ then V is the span of some sequence of vectors. At this point we have not proved this but we will in a later section of this chapter. Before going on to some further results we introduce some more terminology:

Definition 2.24. Assume V is a **subspace of** $\mathbb{R}^n$ and $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is a sequence of **vectors from** $\mathbb{R}^n$ such that $V = Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$. Then we say that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ ***spans*** V or that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is a ***spanning sequence*** for V.

If $V = Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is a **subspace**, $V \neq \{\boldsymbol{0}\}$, then there will be (infinitely) many spanning sequences for V. We will be interested in the question of how to obtain an “efficient” spanning sequence, though for the moment the notion is vague. One aspect of “efficiency” should be that there is no redundancy, that is, we should not be able to remove some vector and still have a spanning sequence. The next theorem begins to get at this.

Theorem 2.3.6. **Reducing spanning sets**
Let $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ be a sequence of **n-vectors** *and set $V = Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$. Assume for some index $j, 1 \leq j \leq k$ that $\boldsymbol{v}_j$ is a* **linear combination** *of the sequence obtained by removing $\boldsymbol{v}_j$. Then $V = Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}, \boldsymbol{v}_{j+1}, \ldots, \boldsymbol{v}_k)$.*

An alternative way to say the same thing is the following:

Theorem 2.3.7. *Let $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_l)$ be a sequence of* **n-vectors**. *Assume the n-vector $\boldsymbol{u}$ is in the* **span** *of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_l)$, that is, $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_l)$. Then $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_l, \boldsymbol{u}) = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_l)$.*

The following example contains all the flavor of a proof.

Example 2.3.3. Let

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{u} = \begin{pmatrix} 2 \\ -1 \\ -2 \\ 1 \end{pmatrix}.$$

Suppose

$$\boldsymbol{w} = \boldsymbol{v}_1 - 2\boldsymbol{v}_2 + 3\boldsymbol{v}_3 - 4\boldsymbol{u} = \begin{pmatrix} -7 \\ 2 \\ 11 \\ -6 \end{pmatrix}.$$

From inspection we have

$$\boldsymbol{u} = 2\boldsymbol{v}_1 - \boldsymbol{v}_2 - 2\boldsymbol{v}_3.$$

We can "plug" this into the expression for $\boldsymbol{w}$ in terms of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u}$ to eliminate $\boldsymbol{u}$:

$$\boldsymbol{w} = \boldsymbol{v}_1 - 2\boldsymbol{v}_2 + 3\boldsymbol{v}_3 - 4(2\boldsymbol{v}_1 - \boldsymbol{v}_2 - 2\boldsymbol{v}_3) = -7\boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 11\boldsymbol{v}_3.$$

More generally, suppose $\boldsymbol{w} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3 + a\boldsymbol{u}$. Then $\boldsymbol{w} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3 + a(2\boldsymbol{v}_1 - \boldsymbol{v}_2 - 2\boldsymbol{v}_3) = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3 + 2a\boldsymbol{v}_1 - a\boldsymbol{v}_2 - 2a\boldsymbol{v}_3 = (a_1 + 2a)\boldsymbol{v}_1 + (a_2 - a)\boldsymbol{v}_2 + (a_3 - 2a)\boldsymbol{v}_4 \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

Thus, we have that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u}) \subseteq Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. However by **Lemma** (2.3.3) we already know that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) \subseteq Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u})$ and therefore we have equality.

Proof. (**Theorem** (2.3.7)). We prove the theorem as alternatively stated as **Theorem** (2.3.7). For the purpose of exposition we take the case that $n = 3$ and $\boldsymbol{u}$ is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$:

$$\boldsymbol{u} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3.$$

From **Lemma** (2.3.3) we know that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) \subseteq Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u})$ so in order to show equality we must show the reverse inclusion, $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u}) \subseteq Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$, that is, every **n-vector** which is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u})$ can be written as a linear combination of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

Assume now that $\boldsymbol{w} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u})$, say

$$\boldsymbol{w} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c\boldsymbol{u}$$

We must show we can write $\boldsymbol{w}$ as a **linear combination** of just $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$:

As in the example we can substitute for $\boldsymbol{u}$:

$$\boldsymbol{w} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c\boldsymbol{u} =$$

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c(a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3) =$$

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + ca_1\boldsymbol{v}_1 + ca_2\boldsymbol{v}_2 + ca_3\boldsymbol{v}_3 =$$

$$(c_1 + ca_1)\boldsymbol{v}_1 + (c_2 + ca_2)\boldsymbol{v}_2 + (c_3 + ca_3)\boldsymbol{v}_3$$

as required. □

This next result, which I have labeled a lemma, will be key to finding "efficient" **spanning sequences** for a **subspace**.

Lemma 2.3.8. *Let $\boldsymbol{u}, \boldsymbol{v}$ be **n-vectors**. Then the following hold:*

1. $Span(\boldsymbol{u}, \boldsymbol{v}) = Span(\boldsymbol{v}, \boldsymbol{u})$.

2. If $c \neq 0$ is a scalar then $Span(\boldsymbol{u}, \boldsymbol{v}) = Span(\boldsymbol{u}, c\boldsymbol{v})$.

3. If c is any scalar then $Span(\boldsymbol{u}, \boldsymbol{v}) = Span(\boldsymbol{u}, c\boldsymbol{u} + \boldsymbol{v})$.

Proof. There is almost no content to the first part which just says that the order in which the **n-vectors** are listed does not effect what vectors are **spanned**. So we move on to parts two and three.

2. Assume $c \neq 0$. To show that $Span(\boldsymbol{u}, \boldsymbol{v}) = Span(\boldsymbol{u}, c\boldsymbol{v})$ we need to prove that any vector which is a **linear combination** of $\boldsymbol{u}$ and $\boldsymbol{v}$ is a linear combination of $\boldsymbol{u}$

and $c\boldsymbol{v}$ and vice versa. Toward that end, suppose we are given $\boldsymbol{w} = a\boldsymbol{u} + b\boldsymbol{v}$. Then $\boldsymbol{w} = a\boldsymbol{u} + \frac{b}{c}(c\boldsymbol{v})$.

Conversely, if $\boldsymbol{w} = a\boldsymbol{u} + b(c\boldsymbol{v})$, an element of $Span(\boldsymbol{u}, c\boldsymbol{v})$, then $\boldsymbol{w} = a\boldsymbol{u} + (bc)\boldsymbol{v}$ which belongs to $Span(\boldsymbol{u}, \boldsymbol{v})$

3. This is similar to part two. Suppose $\boldsymbol{w} = a\boldsymbol{u} + b\boldsymbol{v}$, a **linear combination** of $\boldsymbol{u}$ and $\boldsymbol{v}$. We need to show that $\boldsymbol{w}$ is a linear combination of $\boldsymbol{u}$ and $c\boldsymbol{u} + \boldsymbol{v}$. This is true since $\boldsymbol{w} = (a - bc)\boldsymbol{u} + b(c\boldsymbol{u} + \boldsymbol{v})$.

Conversely, if $\boldsymbol{w} = a\boldsymbol{u} + b(c\boldsymbol{u} + \boldsymbol{v})$ then $\boldsymbol{w} = (a + bc)\boldsymbol{u} + b\boldsymbol{v}$. □

A consequence of **Theorem** (2.3.8) is the following: Suppose we are given some sequence of **n-vectors** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ and we want to find their **span**.

1. If we change the ordering of the sequence this will not change the **span** (since any change in order can be accomplished in steps by interchanging the order of two).

2. If we replace some **n-vector** in the sequence by a non-zero multiple this will not change the **span**.

3. Finally, if we replace some **n-vector** in the sequence by adding a multiple of some vector in the sequence to another vector in the sequence then then **span** is unchanged.

These operations should remind you of something. Namely, **elementary row operations**.

So, starting with a sequence of **n-vectors**, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ if we wish to find a subsequence from amongst them which has the same **span** but cannot be reduced and further by deleting vectors from the sequence then do the following: Turn them on their side and make them the rows of a matrix. Apply **Gaussian elimination** to obtain an **echelon form**. Then the columns obtained (turned on their side) from the **non-zero rows** of this matrix, will have the the same span as the original sequence.

Definition 2.25. ***Row Space of a Matrix, Column Space of a Matrix***

The **span** of the **rows of a matrix** A is called the ***row space*** of the matrix and denoted by $row(A)$. The span of the **columns of a matrix** is called the ***column space*** of the matrix, denoted by $col(A)$.

The next theorem tells us that **row equivalent matrices** have the same **row space**:

Theorem 2.3.9. *If A and A' are* **row equivalent** *(one can be transformed into the other by a sequence of* **elementary row operations***) then $row(A) = row(A')$.*

Warning: The same is not true for the **column space**, namely, **row equivalent matr-**

ices do not necessarily have the same column space.

The next example shows how we can use theory to come to conclusions to determine whether or not a sequence of **n-vectors** is a **spanning sequence** of $\mathbb{R}^n$.

Example 2.3.4. Returning to **Example** (2.3.3), is it possible for the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u})$ to be a **spanning sequence** of $\mathbb{R}^4$?

The answer is no. The argument is as follows: We have seen that $\boldsymbol{u}$ is in the **span** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. This implies that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u}) = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. However, by **Theorem** (2.3.2) a sequence of three **4-vectors** cannot span $\mathbb{R}^4$.

Example 2.3.5. Let

$$\boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Show that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4) = \mathbb{R}^4$.

To demonstrate this we form the matrix $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -2 & 0 & 1 \\ -3 & 0 & 0 & 1 \end{pmatrix}$ and use **Gaussian elimination** to obtain an **echelon form**. One such form is

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & \frac{1}{3} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Every row has a **pivot position**. Therefore by **Theorem** (2.3.1) $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is a **spanning sequence** of $\mathbb{R}^4$.

We conclude this section with an example demonstrating that the **solution set** to a **homogeneous linear system** in n variables is a **span** of a sequence of **n-vectors**.

Example 2.3.6. Find the **solution set** of the following **homogeneous linear system** and express it as the **span** of a sequence of **4-vectors**:

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & x_3 & + & x_4 & = & 0 \\ 2x_1 & + & 3x_2 & + & 3x_3 & + & x_4 & = & 0 \\ 2x_1 & + & 6x_2 & & & + & 4x_4 & = & 0 \\ x_1 & - & x_2 & + & 4x_3 & - & 2x_4 & = & 0 \end{array} \tag{2.11}$$

The **coefficient matrix** of this linear system is

$$A = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 2 & 3 & 3 & 1 \\ 2 & 6 & 0 & 4 \\ 1 & -1 & 4 & -1 \end{pmatrix}$$

After applying **Gaussian elimination** we obtain the **reduced echelon form** , A', of A :

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

The **homogeneous linear system** with coefficient matrix A' is

$$\begin{array}{rcrcrcrcl} x_1 & & & & 3x_3 & - & x_4 & = & 0 \\ & & x_2 & - & x_3 & + & x_4 & = & 0 \end{array} \tag{2.12}$$

Since A and A' are **row equivalent** the **linear systems** in (2.11) and (2.12) are **equivalent**. The linear system (2.12) has two leading variables (x_1, x_2) and two free variables (x_3, x_4). We set x_3 and x_4 equal to parameters: $x_3 = s, x_4 = t$ and express all the variables in terms of s and t :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -3s + t \\ s - t \\ s \\ t \end{pmatrix}$$

We write this as a **sum** of a vector which only involves s and a vector which only involves t

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -3s \\ s \\ s \\ 0 \end{pmatrix} + \begin{pmatrix} t \\ -t \\ 0 \\ t \end{pmatrix}$$

Factoring out the s and t from the respective vectors we get

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = s \begin{pmatrix} -3 \\ 1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

Thus, the typical vector in the solution set of the system (2.12), hence system (2.11), is a **linear combination** of the two vectors $\begin{pmatrix} -3 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$ and, consequently, the solution set is

$$Span\left(\begin{pmatrix} -3 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}\right).$$

What You Can Now Do

1. Given a sequence of **n-vectors** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ and an n-vector $\boldsymbol{u}$ determine if $\boldsymbol{u}$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.

2. Determine if a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of **n-vectors** is a **spanning sequence** of $\mathbb{R}^n$.

3. Express the **solution set** of a **homogeneous linear system** with n variables as a **span** of a sequence of **n-vectors**.

Method (How to do it)

Method 2.3.1. Given a sequence of **n-vectors** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ and an n-vector $\boldsymbol{u}$ determine if $\boldsymbol{u}$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.

Form the **augmented matrix** $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_k \mid \boldsymbol{u})$. If this is the **matrix** of a **consistent linear system** then $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$, otherwise $\boldsymbol{u}$ is not in the span. We determine if the system is consistent by applying **Gaussian elimination** to the augmented matrix to obtain an **echelon form**. Then we determine whether or not the last column of the echelon form is a **pivot column**.

Example 2.3.7. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 4 \\ -3 \\ -2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 7 \\ -6 \\ -3 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ -3 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{u} = \begin{pmatrix} -5 \\ 4 \\ -3 \\ 4 \end{pmatrix}$.

Determine if $\boldsymbol{u}$ is in the **span** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

We form the matrix $\begin{pmatrix} 1 & 2 & 1 & \mid & -5 \\ 4 & 7 & -3 & \mid & 4 \\ -3 & -6 & 1 & \mid & -3 \\ -2 & -3 & 1 & \mid & 4 \end{pmatrix}$ and use **Gaussian elimination** to obtain an **echelon form**:

$$\to \begin{pmatrix} 1 & 2 & 1 & \mid & -5 \\ 0 & -1 & -7 & \mid & 24 \\ 0 & 0 & 4 & \mid & -18 \\ 0 & 1 & 3 & \mid & -6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & \mid & -5 \\ 0 & 1 & 7 & \mid & -24 \\ 0 & 0 & 4 & \mid & -18 \\ 0 & 0 & -4 & \mid & 18 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 1 & \mid & -5 \\ 0 & 1 & 7 & \mid & -24 \\ 0 & 0 & 4 & \mid & -18 \\ 0 & 0 & 0 & \mid & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & \mid & -5 \\ 0 & 1 & 7 & \mid & -24 \\ 0 & 0 & 1 & \mid & -\frac{9}{2} \\ 0 & 0 & 0 & \mid & 0 \end{pmatrix}$$

At this point we see that the last column is not a **pivot column** and consequently that $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

Example 2.3.8. Is the **4-vector** $\boldsymbol{v} = \begin{pmatrix} 1 \\ 3 \\ -2 \\ 0 \end{pmatrix}$ in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$?

Now the relevant matrix is $\begin{pmatrix} 1 & 2 & 1 & \mid & 1 \\ 4 & 7 & -3 & \mid & 3 \\ -3 & -6 & 1 & \mid & -2 \\ -2 & -3 & 1 & \mid & 0 \end{pmatrix}$. As in **Example** (2.3.7) we use **Gaussian elimination** and determine whether or not he last column is a **pivot column**:

$$\to \begin{pmatrix} 1 & 2 & 1 & \mid & 1 \\ 0 & -1 & -7 & \mid & -1 \\ 0 & 0 & 4 & \mid & 4 \\ 0 & 1 & 3 & \mid & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & \mid & 1 \\ 0 & 1 & 7 & \mid & 1 \\ 0 & 0 & 4 & \mid & 1 \\ 0 & 0 & -4 & \mid & 1 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 1 & \mid & 1 \\ 0 & 1 & 7 & \mid & 1 \\ 0 & 0 & 4 & \mid & 1 \\ 0 & 0 & 0 & \mid & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & \mid & 1 \\ 0 & 1 & 7 & \mid & 1 \\ 0 & 0 & 1 & \mid & \frac{1}{4} \\ 0 & 0 & 0 & \mid & 2 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 1 & | & 1 \\ 0 & 1 & 7 & | & 1 \\ 0 & 0 & 1 & | & \frac{1}{4} \\ 0 & 0 & 0 & | & 1 \end{pmatrix}$$

Now we see that the last column is a **pivot column** and therefore $\boldsymbol{v}$ is not in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

Method 2.3.2. Determine if a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of **n-vectors** is a **spanning sequence** of $\mathbb{R}^n$.

Form the matrix $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_k)$ which has the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ as its columns. Use **Gaussian elimination** to obtain an **echelon form** of this matrix and determine whether or not every row has a **pivot position**. If every row has a pivot position then $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is a **spanning sequence** of $\mathbb{R}^n$, otherwise it is not.

Example 2.3.9. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 6 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 2 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 2 \\ -2 \\ -2 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 9 \end{pmatrix}$. Determine if $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k) = \mathbb{R}^n$.

Form the matrix $\begin{pmatrix} 1 & 2 & 3 & 1 \\ 6 & 2 & 2 & 2 \\ 1 & -1 & -2 & 3 \\ -1 & -1 & -2 & 9 \end{pmatrix}$ and use **Gaussian elimination** to obtain an **echelon form**:

$$\to \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & -10 & -16 & -4 \\ 0 & -3 & -5 & 2 \\ 0 & 1 & 1 & 10 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 10 \\ 0 & -3 & -5 & 2 \\ 0 & -10 & -16 & -4 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 10 \\ 0 & 0 & -2 & 32 \\ 0 & 0 & -6 & 96 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 10 \\ 0 & 0 & -2 & 32 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 10 \\ 0 & 0 & 1 & -16 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus we see that the last row does not have a **pivot position** and therefore $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is not a **spanning sequence** of $\mathbb{R}^4$.

Example 2.3.10. If $\boldsymbol{v}_4$ is replaced by $\boldsymbol{v}_4' = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 3 \end{pmatrix}$ is it now the case that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4')$ is a **spanning sequence** of $\mathbb{R}^4$?

Now the relevant matrix is $\begin{pmatrix} 1 & 2 & 3 & 1 \\ 6 & 2 & 2 & 1 \\ 1 & -1 & -2 & 1 \\ -1 & -1 & -2 & 3 \end{pmatrix}$. We apply **Gaussian elimination**:

$$\to \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & -10 & -16 & -5 \\ 0 & -3 & -5 & 0 \\ 0 & 1 & 1 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & -3 & -5 & 0 \\ 0 & -10 & -16 & -5 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & -2 & 12 \\ 0 & 0 & -6 & 35 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & -6 & 35 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Now every row has a **pivot position** and consequently $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4')$ is a **spanning sequence** of $\mathbb{R}^4$.

Method 2.3.3. Express the **solution set** of a **homogeneous linear system** with n variables as a **span** of a sequence of **n-vectors**

When we use **Method** (1.2.4) to solve a **homogeneous linear system** in n variables the **general solution** can be expressed in the form of a **sum** of **n-vectors** times parameters, where the parameters are free to take any real value. Consequently, all **linear combinations** of these vectors are solutions. Thus, the **solution set** to the homogeneous system is the **span** of these vectors.

Example 2.3.11. Express the **solution set** to the following **homogeneous system** as a **span** of vectors:

$$\begin{array}{rcrcrcrcl} 3x_1 & - & 4x_2 & + & 6x_3 & + & 5x_4 & = & 0 \\ 2x_1 & - & 3x_2 & + & 4x_3 & + & 4x_4 & = & 0 \\ 2x_1 & - & x_2 & + & 4x_3 & & & = & 0 \end{array} \tag{2.13}$$

The **coefficient matrix** of this **homogeneous linear system** is $\begin{pmatrix} 3 & -4 & 6 & 5 \\ 2 & -3 & 4 & 4 \\ 2 & -1 & 4 & 0 \end{pmatrix}$. We use **Gaussian elimination** we obtain the **reduced echelon form** for this matrix:

$$\to \begin{pmatrix} 1 & -\frac{4}{3} & 2 & \frac{5}{3} \\ 2 & -3 & 4 & 4 \\ 2 & -1 & 4 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & -\frac{4}{3} & 2 & \frac{5}{3} \\ 0 & -\frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{5}{3} & 0 & -\frac{10}{3} \end{pmatrix} \to \begin{pmatrix} 1 & -\frac{4}{3} & 2 & \frac{5}{3} \\ 0 & 1 & 0 & -2 \\ 0 & \frac{5}{3} & 0 & -\frac{10}{3} \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & -\frac{4}{3} & 2 & \frac{5}{3} \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The **homogeneous linear system** with this matrix equal to its coefficient matrix is

$$\begin{array}{rcrcrcrcl} x_1 & & & + & 2x_3 & - & x_4 & = & 0 \\ & & x_2 & & & - & 2x_4 & = & 0 \end{array} \tag{2.14}$$

This system is **equivalent** to the original system and therefore has an identical solution set.

The **leading variables** are x_1, x_2 and the **free variables** are x_3, x_4. We set the free variables equal to parameters, $x_3 = s, x_4 = t$ and obtain

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

Thus, the **solution set** to the **linear system** in (2.14), hence linear system (2.13), is $Span\left(\begin{pmatrix} -2 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}\right)$.

Exercises

Let

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 5 \\ 4 \\ 2 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}.$$

In exercises 1-4 determine if the given **3-vector** $\boldsymbol{v}$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. See **Method** (2.3.1)

1. $\boldsymbol{v} = \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix}$

2. $\boldsymbol{v} = \begin{pmatrix} 4 \\ 1 \\ -5 \end{pmatrix}$

3. $\boldsymbol{v} = \begin{pmatrix} 1 \\ 2 \\ -7 \end{pmatrix}$

4. $\boldsymbol{v} = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}$

Let

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -2 \end{pmatrix}.$$

Set $S = Span(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$.

In exercises 4-8 determine if the given vector $\boldsymbol{u} \in S$. See **Method** (2.3.1)

5. $\boldsymbol{u} = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$

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

7. $\boldsymbol{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

8. $\boldsymbol{u} = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

In exercises 9 - 18 determine if the columns of the given matrix A span the appropriate $\mathbb{R}^n$. If you can give an explanation without computing then do so. See **Method** (2.3.2)

9. $\begin{pmatrix} 2 & 3 & 4 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

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

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

12. $\begin{pmatrix} 2 & 1 & 3 & 3 \\ 2 & 2 & 4 & 2 \\ 1 & 2 & 3 & -1 \end{pmatrix}$

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

14. $\begin{pmatrix} 1 & 3 & 2 & 3 \\ 1 & 2 & 1 & -4 \\ 1 & 1 & 2 & -3 \\ 1 & 2 & 3 & 4 \end{pmatrix}$

15. $\begin{pmatrix} 1 & 1 & 0 & 1 \\ 3 & 2 & -1 & 1 \\ 3 & 5 & 4 & 1 \\ 2 & 3 & 2 & 1 \end{pmatrix}$

16. $\begin{pmatrix} 2 & -1 & 0 & 1 \\ 5 & -2 & 0 & 2 \\ 0 & 1 & 0 & 1 \end{pmatrix}$

17. $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 5 & -2 \\ 2 & 3 & -4 \end{pmatrix}$

18. $\begin{pmatrix} 2 & 1 & 3 & 5 \\ 3 & 2 & -4 & 1 \\ 5 & 3 & -2 & 5 \\ 2 & 1 & 2 & 4 \end{pmatrix}$

In 19-22 express the **solution set** to the given **homogeneous linear system** as a **span of a sequence of vectors**. See **Method** (2.3.3)

19.

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

20.

$$\begin{array}{ccccccccc} x_1 & - & x_2 & + & 2x_3 & - & 2x_4 & = & 0 \\ 2x_1 & - & 3x_2 & & & + & x_4 & = & 0 \\ 3x_1 & - & 2x_2 & + & 9x_3 & - & 10x_4 & = & 0 \end{array}$$

21.

$$\begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & + & 3x_4 & = & 0 \\ 2x_1 & + & 3x_3 & - & 2x_3 & + & 3x_4 & = & 0 \\ x_1 & + & 3x_2 & + & 5x_3 & + & 9x_4 & = & 0 \\ x_1 & + & 3x_2 & - & 7x_3 & - & 3x_4 & = & 0 \end{array}$$

22.

$$\begin{array}{ccccccccccc} x_1 & - & x_2 & + & 2x_3 & - & 2x_4 & + & 2x_5 & = & 0 \\ -x_1 & + & 2x_2 & - & 3x_3 & - & 2x_4 & - & 3x_5 & = & 0 \\ 2x_1 & - & 3x_2 & + & 4x_3 & + & 3x_4 & + & 3x_5 & = & 0 \\ & & x_2 & - & x_3 & & & - & x_5 & = & 0 \\ x_1 & - & x_2 & + & x_3 & + & x_4 & + & 2x_5 & = & 0 \end{array}$$

In exercises 23 - 28 answer true or false and give an explanation.

23. If S is a sequence of twenty **4-vectors** then S will always **span** $\mathbb{R}^4$.

24. If S is a sequence of twenty **3-vectors** from then S will sometimes **span** $\mathbb{R}^3$.

25. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is a **spanning sequence** of $\mathbb{R}^4$ then also $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3 + \boldsymbol{v}_4)$ is a spanning sequence of $\mathbb{R}^4$.

26. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is a **spanning sequence** of $\mathbb{R}^4$ then also $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4, \boldsymbol{v}_3)$ is a spanning sequence of $\mathbb{R}^4$.

27. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is a **spanning sequence** of $\mathbb{R}^4$ then $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a spanning sequence of $\mathbb{R}^3$. $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = \mathbb{R}^3$.

28. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is a sequence of **4-vectors** and $\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3 + \boldsymbol{v}_4 = \boldsymbol{0}_4$ then $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is not a **spanning sequence** .

Challenge Exercises (Problems)

1. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \in \mathbb{R}^3$ and assume that

$\boldsymbol{v}_3 = 2\boldsymbol{v}_1 - 5\boldsymbol{v}_2$.

a) Explain why the **linear system** with **augmented matrix** $\hat{A} = (\boldsymbol{v}_1 \ \boldsymbol{v}_2 \mid \boldsymbol{v}_3)$ is **consistent**.

b) Why can you conclude that the last column of $\hat{A}$ is not a **pivot column**.

c) Explain why $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **not** a **spanning sequence** of $\mathbb{R}^3$.

2. Assume that $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are **n-vectors** and $\boldsymbol{v}_n \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{n-1})$. Give an explanation for why $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) \neq \mathbb{R}^n$.

3. Assume that $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are **4-vectors** and let A be the matrix with these vectors as columns:

$A = (\boldsymbol{v}_1\boldsymbol{v}_2\boldsymbol{v}_3)$ and assume that each column is a **pivot column**.

a) Let $\boldsymbol{v}_4$ be a **4-vector** and assume that $\boldsymbol{v}_4 \notin Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Explain why the **linear system** with **augmented matrix** $\hat{A} = \left(\boldsymbol{v}_1\boldsymbol{v}_2\boldsymbol{v}_3|\boldsymbol{v}_4\right)$ is **inconsistent**.

b) Explain why $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ $=\mathbb{R}^4$.

4. Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ be a sequence of **4-vectors** and assume that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ = $\mathbb{R}^4$. Prove that $\boldsymbol{v}_4 \notin Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

5. Let U and V be **subspaces** of $\mathbb{R}^n$. Prove that $U \cap V$, consisting of all vectors which belong to both U and V, is a subspace of $\mathbb{R}^n$.

6. Let U and V be **subspaces** of $\mathbb{R}^n$. Let $U + V$ consist of all vectors of the form $\boldsymbol{u} + \boldsymbol{v}$ where $\boldsymbol{u} \in U, \boldsymbol{v} \in V$ Prove that $U + V$ is a subspace of $\mathbb{R}^n$.

Quiz Solutions

1.Yes, $\begin{pmatrix} 2 \\ 4 \\ 7 \\ 2 \end{pmatrix} = \boldsymbol{u}_1 + \boldsymbol{u}_2 - \boldsymbol{u}_3$.

Not right, see **Method** (2.2.1) and **Method** (2.2.2)

2. $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 27\boldsymbol{v}_1 - 7\boldsymbol{v}_2 - 4\boldsymbol{v}_3$

$\boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = -12\boldsymbol{v}_1 + 3\boldsymbol{v}_2 + 2\boldsymbol{v}_3$

$\boldsymbol{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -5\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3.$

Not right, see **Method** (2.2.1) and **Method** (2.2.2)

3. $\boldsymbol{u} = \begin{pmatrix} -3 \\ 2 \\ 5 \end{pmatrix} = -3\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} =$

$(-3)(27\boldsymbol{u}_1 - 7\boldsymbol{u}_2 - 4\boldsymbol{u}_3) + 2(-12\boldsymbol{u}_1 + 3\boldsymbol{u}_2 + 2\boldsymbol{u}_3) + 5(-5\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3) = -130\boldsymbol{u}_1 + 32\boldsymbol{u}_2 + 21\boldsymbol{u}_3.$

Not right, see **Method** (2.2.1) and **Method** (2.2.2)

2.4. Linear independence in $\mathbb{R}^n$

Am I Ready for This Material

In order to follow this section and master its ideas and methods you need to be familiar with many previously defined concepts. **Those that are used extensively in this section are**:

linear system

solution of a linear system

solution set of a linear system

consistent linear system

inconsistent linear system

equivalent linear systems

echelon form of a linear system

leading variable

free variable

matrix

augmented matrix of a linear system

row equivalence of matrices

matrix in row echelon form

matrix in reduced row echelon form

echelon form of a matrix

reduced echelon from of a matrix

pivot positions of a matrix

pivot columns of a matrix

an $n-$vector

equality of $n-$vectors

$n-$space, $\mathbb{R}^n$

adddition of $n-$vectors

scalar multiplication of an $n-$vector

negative of a vector

the zero vector

the standard basis for $\mathbb{R}^n$

linear combination of vectors

span of a sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ from $\mathbb{R}^n$

spanning sequence of $\mathbb{R}^n$

subspace of $\mathbb{R}^n$

spanning sequence of a subspace V of $\mathbb{R}^n$

You will also need to be able to execute several procedures previously described. **The following are algorithms that we will use in this section**:

Gaussian elimination

procedure for solving a linear system

Quiz

Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 4 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 1 \end{pmatrix}$, and set $A = \begin{pmatrix} 1 & 2 & 1 & 3 \\ 1 & 3 & 0 & 4 \\ 1 & 1 & 2 & 2 \\ 1 & -1 & 4 & 1 \end{pmatrix}$, the matrix wtih columns the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ of **4-vectors**.

1. Compute the **reduced echelon form** of the matrix A. Use this to determine the **pivot positions** and **pivot columns** of A?

2. Is $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ a **spanning sequence** of $\mathbb{R}^4$?

3. Assume that A is the **coefficient matrix** of a **homogeneous linear system**. The number of **solutions** of this system is

a) Zero, the system is **inconsistent**
b) 1
c) 2
d) infinite
d) Cannot be determined.

4. Assume that A is the **augmented matrix** of an **inhomogeneous linear system**. The number of **solutions** of this system is
a) Zero, the system is **inconsistent**
b) 1
c) 2
d) infinite
d) cannot be determined.

5. Let $B = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 4 \\ 1 & 0 & 2 & -1 \end{pmatrix}$ and assume B is the **coefficient matrix** of a **linear system** . Determine the **solution set** of this system.

6. Let $\bar{b} = \boldsymbol{v}_1 - 3\boldsymbol{v}_2 + 2\boldsymbol{v}_4$. The number of **solutions** of the **inhomogeneous linear system** with **matrix** $[A|\bar{b}]$ is
a) Zero, the system is **inconsistent**
b) 1
c) 2
d) infinite
e) cannot be determined.

Quiz Solutions

New Concepts

This section introduces several new concepts. Two are especially important and fundamental to all the material of the course. Hereafter, these terms will be used in virtually every subsequent section. There concepts are:

linearly dependent sequence of n-vectors

linearly independent sequence of n-vectors

To simplify the formulation of these definitions we will also introduce the following concepts:

dependence relation on a sequence of n-vectors

trivial and non-trivial dependence relation on a sequence of n-vectors

Theory (Why it Works)

Warning: The material introduced here is fundamental and will be used in nearly every subsequent section of the book. We begin with a definition:

Definition 2.26. Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be a sequence of **n-vectors**. By a ***dependence relation*** on $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ we mean any **linear combination** $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$ which is equal to the **zero vector**.

There is always an obvious **dependence relation** on a sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$, namely the **linear combination** all of whose components are zero: $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_k = \boldsymbol{0}_n$. We single out this dependence relation and give it a special designation.

Definition 2.27. Let $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ be a sequence of **n-vectors**. The particular **dependence relation** $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_k = \boldsymbol{0}_n$ is called the ***trivial dependence relation***. Any other dependence relation is called a ***non-trivial dependence relation***.

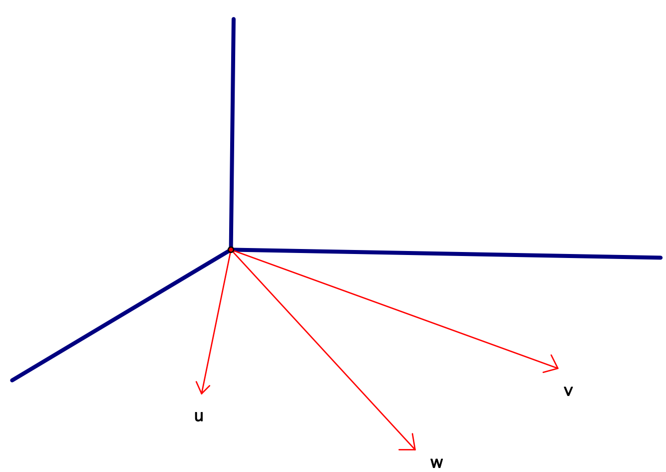


Figure 2.4.1: Three dependent vectors in $\mathbb{R}^3$

We now come to two of the most fundamental and important concepts of the entire subject of linear algebra:

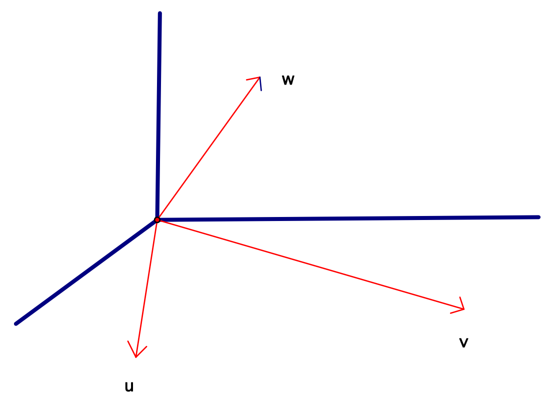


Figure 2.4.2: Three independent vectors in $\mathbb{R}^3$

Definition 2.28. If there exists a **non-trivial dependence relation** for the sequence of **n-vectors** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ then we say that the sequence is ***linearly dependent***.

On the other hand, if the only **dependence relation** for the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is the **trivial dependence relation** then we say that the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is ***linearly independent***

Linear dependence is illustrated in Figure (2.4.1) which shows three **vectors** contained in the xy-plane $\mathbb{R}^3$ (they all have z-coordinate equal to zero).

On the other hand, Figure (2.4.2) shows three vectors in $\mathbb{R}^3$ which do not belong to any common plane. These vectors are **linearly independent**.

For the purposes of computation, alternative formulations of dependence and independence are required: A sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ of **n-vectors** is **linearly dependent** if and only if there exists scalars $c_1, \dots, c_k$, not all zero (**at least one is non-zero**) such that $c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}_n$.

On the other hand, the sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ of **n-vectors** is **linearly independent** if and only if whenever $c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}_n$ then $c_1 = \cdots = c_k = 0$.

We will determine criteria for deciding whether or not a sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ of **n-vectors** is **linearly dependent** or not. In the course of this, we will demonstrate how to find a **non-trivial dependence relation** when a sequence of vectors is linearly dependent. As we will see, this will reduce to finding **solutions** to a **homogeneous linear system** and, as usual these questions can be answered using **Gaussian elimination** to obtain the **reduced echelon form** of a suitable matrix. Before we get to the formal treatment we do an example.

Example 2.4.1. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -3 \\ 2 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ -1 \\ -1 \\ -1 \end{pmatrix}$. Determine if the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly dependent or linearly independent**.

Using the **alternative formulation**, we have to determine if there are scalars, c_1, c_2, c_3, not all zero, such that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = 0$. Substituting the given **4-vectors** yields:

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

After performing the vector operations (**addition** and **scalar multiplication**) this yields the following vector equation:

$$\begin{pmatrix} c_1 + 3c_3 \\ -2c_1 + c_2 - c_3 \\ c_1 - 3c_2 - c_3 \\ 2c_2 - c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

By the definition of **equality of vectors** this is equivalent to the **homogeneous linear system**:

$$\begin{array}{rcrcrcl} c_1 & & & + & 3c_3 & = & 0 \\ -2c_1 & + & c_2 & - & c_3 & = & 0 \\ c_1 & - & 3c_2 & - & c_3 & = & 0 \\ & & 2c_2 & - & c_3 & = & 0 \end{array}$$

Thus, the original question - is the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ **linearly dependent or linearly independent** - reduces to the question of whether or not there are **non-trivial solutions** to the **homogeneous linear system** with **coefficient matrix**

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

that is, the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3)$. This matrix has **reduced echelon form**

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

Thus, the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**.

Example 2.4.2. Let $\boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$. Now we determine if sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent**.

We have to determine if there are scalars c_1, c_2, c_3, c_4, not all zero, such that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 = \mathbf{0}$. Substituting the given vectors we obtain:

$$c_1 \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ -3 \\ 2 \end{pmatrix} + c_3 \begin{pmatrix} 3 \\ -1 \\ -1 \\ -1 \end{pmatrix} + c_4 \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

This yields the vector equation

$$\begin{pmatrix} c_1 + 3c_3 + c_4 \\ -2c_1 + c_2 - c_3 \\ c_1 - 3c_2 - c_3 \\ 2c_2 - c_3 - c_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

which is the same as the **homogeneous linear system** :

$$\begin{array}{rcrcrcrcl} c_1 & & & + & 3c_3 & + & c_4 & = & 0 \\ -2c_1 & + & c_2 & - & c_3 & & & = & 0 \\ c_1 & - & 3c_2 & - & c_3 & & & = & 0 \\ & & 2c_2 & - & c_3 & - & c_4 & = & 0 \end{array}.$$

As in **Example** (2.4.1) this comes down to whether the **linear system** has **free variable** equivalently, if there are **non-pivot columns** in the **coefficient matrix**

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

After applying **Gaussian elimination** we obtain the **echelon form**

$$\begin{pmatrix} 1 & 0 & 3 & 1 \\ 0 & 1 & 5 & 2 \\ 0 & 0 & 1 & \frac{5}{11} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since there are **non-pivot columns**, the **homogeneous linear system** with **coefficient matrix** A has **non-trivial solutions**. Consequently, the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent**.

Continuing with the use of **Gaussian elimination** , we obtain the **reduced echelon form** in order to find a **non-trivial dependence relation**:

$$A' = \begin{pmatrix} 1 & 0 & 0 & -\frac{4}{11} \\ 0 & 1 & 0 & -\frac{3}{11} \\ 0 & 0 & 1 & \frac{5}{11} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The **homogeneous linear system** with **coefficient matrix** A' is

$$\begin{array}{ccccccc} c_1 & & & - & \frac{4}{11}c_4 & = & 0 \\ & c_2 & & - & \frac{3}{11}c_4 & = & 0 \\ & & c_3 & + & \frac{5}{11}c_4 & = & 0 \end{array}$$

There are three **leading variable** (c_1, c_2, c_3) and one **free variable** (c_4). Setting $c_4 = t$ we get

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} \frac{4}{11}t \\ \frac{3}{11}t \\ -\frac{5}{11}t \\ t \end{pmatrix}.$$

For each value of t we get a **dependence relation** . For example, taking $t = 11$ (chosen to "clear fractions") we get the dependence relation $4\boldsymbol{v}_1 + 3\boldsymbol{v}_2 - 5\boldsymbol{v}_3 + 11\boldsymbol{v}_4 = \boldsymbol{0}$.

Example (2.4.1) and and **Example** (2.4.2) suggest the following general result:

Theorem 2.4.1. *Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ be a sequence of* **n-vectors** *. Then the following are equivalent:*

1. $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is **linearly independent***;*

2. The **homogeneous linear system** *with* **coefficient matrix** *$(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_k)$ has only the* **trivial solution***;*

3. The matrix $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_k)$ with columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ has k **pivot columns** *, that is, each column of A is a pivot column.*

Assume $k > n$. Then the matrix A in **Theorem** (2.4.1) is $n \times k$ and therefore has more columns than rows. Since each row contains at most one **pivot position** there are at most n pivot positions. Therefore it is not possible for each column to be a **pivot column** . Consequently, when $k > n$ the set of columns is necessarily **linearly dependent**. We summarize this as a formal result:

Theorem 2.4.2. *If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is a* **linearly independent** *sequence of* **n-vectors** *then $k \leq n$.*

This is equivalent to the following sufficient condition for a sequence of **n-vectors** to be **linearly dependent**:

Theorem 2.4.3. *Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ be a sequence of* **n-vectors** *. If $k > n$ then the sequence is* **linearly dependent***.*

The test for a sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ of **n-vectors** to be **linearly independent** or **linearly dependent** is a lot like the one used to determine if the sequence is a **spanning sequence** for $\mathbb{R}^n$: To test for linear independence or linear dependence you decide if every **column** of the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_k)$ is a **pivot column** . In the test to determine if the sequence spans $\mathbb{R}^n$ you must determine if every row of the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_k)$ contains a **pivot position** .

Don't mix these up; in general they are not the same thing. However, in one situation they are equivalent: When we have a sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ of length n consisting of **n-vectors** . In this situation the matrix $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_n)$ has equally many rows as columns (and we say that A is a ***square matrix***). Consequently every row has a **pivot position** if and only if every column is **pivot column** . This proves the following result which we refer to as the "Half is Good Enough Theorem".

Theorem 2.4.4. *Let $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_n)$ be an $n \times n$ matrix so that the columns are* **n-vectors** *. Then the following are equivalent:*

i. Every row of A has a **pivot position** *;*

ii. The sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ is a **spanning sequence** *;*

iii. Every column of A is a **pivot column** *;*

iv. The sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ is **linearly independent***.*

An explanation is in order for why we have referred to this as the "half is good enough" theorem. A **basis** of $\mathbb{R}^n$ is a sequence of vectors $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ which is **linearly independent** and **spans** $\mathbb{R}^n$. In order to span we must have $k \geq n$ from Section 2.3. We have just seen in order to be **linearly independent** we must have $k \leq n$. Therefore, it must be the case that $k = n$. Theorem (2.4.4) states that we can conclude that a sequence of length n consisting of **n-vectors** is a basis provided we know *either* the sequence is linearly independent or that it spans $\mathbb{R}^n$. That the other property is satisfied is a consequence of the theorem. Thus, half the necessary information is good enough.

We next investigate some consequences that we can draw from the information that a sequence of **n-vectors** is **linearly dependent**. We begin with the simplest case, that of a single n-vector.

Thus, assume that $\boldsymbol{v}$ is **linearly dependent**. Then there is a scalar $c \neq 0$ $c\boldsymbol{v} = \mathbf{0}_n$. But then we may divide by c (multiply by $\frac{1}{c}$) to obtain $\boldsymbol{v} = \frac{1}{c}\mathbf{0}_n = \mathbf{0_n}$ and therefore $\boldsymbol{v}$ is the **zero vector**.

Conversely, suppose $\boldsymbol{v} = \mathbf{0}_n$, the zero vector. Then $1\boldsymbol{v} = 1\mathbf{0}_n = \mathbf{0}_n$ and so, since $1 \neq 0$, the sequence $(\mathbf{0}_n)$ is **linearly dependent**.

We have therefore shown:

Theorem 2.4.5. *A sequence consisting of a single* **n-vector** *, $(\boldsymbol{v})$, is* **linearly dependent** *if and only if $\boldsymbol{v} = \mathbf{0}_n$.*

We next investigate what can be said when a sequence of two **n-vectors** is **linearly dependent** . Our goal is to prove the following:

Theorem 2.4.6. *Let $(\boldsymbol{u}, \boldsymbol{v})$ be a sequence of two non-zero* **n-vectors** *. Then the following are equivalent:*

1. $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly dependent***;*

2. $\boldsymbol{u}$ and $\boldsymbol{v}$ are **scalar multiples** *of one another.*

]

Proof. Suppose $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly dependent**. Then there are scalars α, β, not both zero, such that $\alpha\boldsymbol{u} + \beta\boldsymbol{v} = \boldsymbol{0}_n$. Suppose $\alpha = 0$. Then $\beta \neq 0$ and then $(\boldsymbol{v})$ is linearly dependent which implies that $\boldsymbol{v} = \boldsymbol{0}_n$ by **Theorem** (2.4.5), contrary to assumption. We therefore conclude that $\alpha \neq 0$ and in a similar fashion that $\beta \neq 0$.

Because $\alpha\beta \neq 0$ we can solve for either vector in terms of the other: $\boldsymbol{u} = -\frac{\beta}{\alpha}\boldsymbol{v}, \boldsymbol{v} = -\frac{\alpha}{\beta}\boldsymbol{u}$.

Conversely, suppose one of the vectors is a **scalar multiple** of the other, say $\boldsymbol{u} = \gamma\boldsymbol{v}$. It follows from this equality that $(1)\boldsymbol{u} + (-\gamma)\boldsymbol{v} = 0$ and therefore $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly dependent**. $\square$

We now look at the general case. The following result will be very useful for theoretic purposes later on in the book:

Theorem 2.4.7. *Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ be a sequence of* **n-vectors** *with $k \geq 2$. Then the following hold:*

1. S is **linearly dependent** *if and only if some vector of S is a* **linear combination** *of the remaining vectors.*

2. Assume for some i with $1 \leq i < k$ that $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_i)$ is **linearly independent**. *Then S is* **linearly dependent** *if and only if there is a $j > i$ such that $\boldsymbol{v}_j$ is a* **linear combination** *of the sequence of vectors which proceed it: $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{j-1})$.*

Proof. 1. We illustrate and sketch the proof for the case $k = 4$, the general case is proved in exactly the same way. Let us assume the sequence $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent** and prove that some $\boldsymbol{v}_j$ is a **linear combination** of the others. Because the sequence S is dependent there exists scalars c_1, c_2, c_3, c_4, not all of them zero, such that

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 = \boldsymbol{0}.$$

By relabeling, if necessary, we can assume that $c_4 \neq 0$. By subtracting $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3$ from both sides we obtain:

$$c_4\boldsymbol{v}_4 = -(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3) = -c_1\boldsymbol{v}_1 - c_2\boldsymbol{v}_2 - c_3\boldsymbol{v}_3.$$

Multiplying by $\frac{1}{c_4}$, which is permissable since $c_4 \neq 0$, we obtain

$$\boldsymbol{v}_4 = (-\frac{c_1}{c_4})\boldsymbol{v}_1 + (-\frac{c_2}{c_4})\boldsymbol{v}_2 + (-\frac{c_3}{c_4})\boldsymbol{v}_3$$

an expression of $\boldsymbol{v}_4$ as a linear combination of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

Conversely, suppose one of the **n-vectors** $\boldsymbol{v}_j$, e.g. $\boldsymbol{v}_4$ is a **linear combination** of the remaining vectors, say

$$\boldsymbol{v}_4 = \alpha\boldsymbol{v}_1 + \beta\boldsymbol{v}_2 + \gamma\boldsymbol{v}_3.$$

We then get the identity

$$(-\alpha)\boldsymbol{v}_1 + (-\beta)\boldsymbol{v}_2 + (-\gamma)\boldsymbol{v}_3 + (1)\boldsymbol{v}_4 = \mathbf{0}$$

and so $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent**.

2. We now go on to the second part and prove it in full generality. Since the **n-vectors** are **linearly dependent** there is a **non-trivial depedence relation**, that is, an expression

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}.$$

where at least one $c_l \neq 0$.

Let j be the largest index such that $c_j \neq 0$. Then by dropping the succeeding terms whose coefficients are equal to zero the relation becomes

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + \cdots + c_j\boldsymbol{v}_j = \mathbf{0}. \tag{2.15}$$

We claim that it must be the case that $j > i$. Assume to the contrary that $j \leq i$. If $j = i$ then the dependence relation (2.15) becomes

$$c_1\boldsymbol{v}_1 + \cdots + c_i\boldsymbol{v}_i = \mathbf{0}_n$$

which implies that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i)$ is **linearly dependent** contrary to the hypothesis. On the other hand, if $j < i$ then the dependence relation (2.15) becomes

$$c_1\boldsymbol{v}_1 + \cdots + c_i\boldsymbol{v}_j + 0\boldsymbol{v}_{j+1} + \cdots + c_i\boldsymbol{v}_i = \mathbf{0}_n$$

which again implies that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i)$ is **linearly dependent**, a contradiction. Thus, $j > i$ as required.

Now we can solve for $\boldsymbol{v}_j$ in terms of the preceeding vectors:

$$c_j\boldsymbol{v}_j = -(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + \cdots + c_{j-1}\boldsymbol{v}_{j-1}) =$$
$$-c_1\boldsymbol{v}_1 - c_2\boldsymbol{v}_2 - c_3\boldsymbol{v}_3 - \cdots - c_{j-1}\boldsymbol{v}_{j-1}$$

Dividing by c_j we get

$$\boldsymbol{v}_j = (-\frac{c_1}{c_j})\boldsymbol{v}_1 + (-\frac{c_2}{c_j})\boldsymbol{v}_2 + \cdots + (-\frac{c_{j-1}}{c_j})\boldsymbol{v}_{j-1}.$$

The converse follows in the same way as in the first part. $\square$

We now summarize and collect some useful facts about **linear dependence** and **linear independence** and state these as a theorem.

Theorem 2.4.8. *1. Any finite sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of* **n-vectors** *which contains the* **zero vector** *, $\mathbf{0}_n$, is* **linearly dependent***.*

2. Any finite sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of **n-vectors** *which has repeated vectors is* **linearly dependent***.*

3. Any finite sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of **n-vectors** *which contains a* **linearly dependent** *subsequence is itself linearly dependent. Another way of saying this is that a sequence obtained by adding one or more n-vectors to a linearly dependent sequence is itself linearly dependent.*

The last statement is logically equivalent to the following result:

Theorem 2.4.9. *Let $S = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ be an* **linearly independent** *sequence of* **n-vectors***. Any sequence obtained from S by deleting a vector is linearly independent.*

Proof. Rather than write out long, formal proofs we demonstrate by examples which fully illustrate the rigorous argument:

1. In $\mathbb{R}^3$ consider the sequence $(\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2)$ where $\boldsymbol{e}_1, \boldsymbol{e}_2$ are the first two vectors of the **standard basis of $\mathbb{R}^3$**:

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

It is simple to write a **non-trivial linear dependence relation** on these vectors, namely,

$$(1)\mathbf{0}_3 + 0\boldsymbol{e}_1 + 0\boldsymbol{e}_2 = \mathbf{0}_3.$$

In the general case, multiply the **zero vector** , $\mathbf{0}_n$, by 1 and all the other vectors by the scalar 0 and this will give a **non-trivial dependence relation**.

2. Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ where

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

It is simple to write a **non-trivial dependence relation** for this sequence, namely

$$(1)\boldsymbol{v}_1 + (0)\boldsymbol{v}_2 + (-1)\boldsymbol{v}_3 =$$

$$(1)\begin{pmatrix}1\\1\\1\end{pmatrix} + (0)\begin{pmatrix}1\\0\\0\end{pmatrix} + (-1)\begin{pmatrix}1\\1\\1\end{pmatrix} =$$

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

In general, multiply the first occurence of a repeated vector by the scalar one, the second occurence of that vector by the scalar -1 and all other vectors by the scalar 0. This will give a **non-trivial dependence relation**.

3. The sequence $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ where

$$\boldsymbol{v}_1 = \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix}0\\0\\1\\-1\end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix}2\\-2\\3\\-3\end{pmatrix}$$

is **linearly dependent** since

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

If we add $\boldsymbol{v}_4 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}$ to obtain the sequence $S' = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ we obtain a **linearly dependent** sequence since we can extend the above dependence relation by multiplying $\boldsymbol{v}_4$ by the scalar zero:

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

Theorem (2.4.9) is logically equivalent to part three of the **Theorem** (2.4.8): If $S = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is a **linearly independent** sequence of vectors and S^* is obtained by taking away some vector, $\boldsymbol{v}_j$, then either S^* is linearly independent or **linearly dependent**.

Suppose that S^* is linearly dependent, contrary to what we are trying to show. Now S can be obtained from S^* by adding $\boldsymbol{v}_j$. But if S^* is linearly dependent then by part three of **Theorem** (2.4.8), S is linearly dependent, contrary to hypothesis. So it must be that S^* is linearly independent. $\square$

We conclude with two more theoretical results. Both are exceedingly important for later developments.

Theorem 2.4.10. *Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be a sequence of* **n-vectors** .

1. If S is **linearly independent** *then any vector $\boldsymbol{v}$ in the* **span** *of S is expressible in one and only one way as a* **linear combination** *of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.*

2. If every vector in $Span(S)$ can be written as a **linear combination** *of the vectors in S in only one way then S is* **linearly independent**

Proof. 1. As in previous proofs, for the sake of exposition, we use an actual value for k, say $k = 3$ rather than write out the general proof. Suppose

$$\boldsymbol{v} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3 = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + b_3\boldsymbol{v}_3.$$

We need to show that $a_1 = b_1, a_2 = b_2, a_3 = b_3$.

Since these two vectors are **equal** we can subtract to get the **zero vector** , $\mathbf{0}_n$:

$$(a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3) - (b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + b_3\boldsymbol{v}_3) = \mathbf{0}_n.$$

After performing the vector operations of distributing, commuting and regrouping we get:

$$(a_1 - b_1)\boldsymbol{v}_1 + (a_2 - b_2)\boldsymbol{v}_2 + (a_3 - b_3)\boldsymbol{v}_3 = 0.$$

This is a **dependence relation** on $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ which is **linearly independent**. Therefore by **alternate formulation of linear independence** we must have $a_1 - b_1 = a_2 - b_2 = a_3 - b_3 = 0$ which implies that $a_1 = b_1, a_2 = b_2, a_3 = b_3$, as required.

2. If every vector in $Span(S)$ is a uniquely a **linear combination** of the vectors in S then, in particular, this applies to the **zero vector** , $\mathbf{0}_n$. However, $\mathbf{0}_n = 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_k$. Since we are assuming there is a unique such expression this must be it. Consequently, the only **dependence relation** is the **trivial dependence relation** and this implies that S is **linearly independent**. $\square$

We now come to our final result which gives criteria under which the expansion of a **linearly independent** sequence is linearly independent.

Theorem 2.4.11. *Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be a* **linearly independent** *sequence of* **n-vectors** *. If $\boldsymbol{u}$ is not in Span S then the sequence $S^* = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{u})$ obtained by adjoining $\boldsymbol{u}$ to S is linearly independent.*

Proof. For the purposes of exposition, let's assume that $k = 4$. We do a **proof by contradiction**, that is, we assume the conclusion is false and derive a contradiction.

Suppose to the contrary that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{u})$ is a **linearly dependent** sequence of vectors. Then by **alternate formulation of linear dependence** there are scalars c_1, c_2, c_3, c_4, d, not all of them zero, such that

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 + d\boldsymbol{u} = \boldsymbol{0}_n.$$

Suppose first that $d = 0$. Then one of c_1, c_2, c_3, c_4 is not zero and $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 = \boldsymbol{0}_n$ and so is a **non-trivial dependence relation** on S. Therefore S is **linearly dependent**
which contradict one of our hypotheses.

Thus, we may assume that $d \neq 0$. But then, as we have previously seen, this implies that $\boldsymbol{u}$ is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$:

$$\boldsymbol{u} = (-\frac{c_1}{d})\boldsymbol{v}_1 + (-\frac{c_2}{d})\boldsymbol{v}_2 + (-\frac{c_3}{d})\boldsymbol{v}_3 + (-\frac{c_4}{d})\boldsymbol{v}_4.$$

What is wrong with this? It means that $\boldsymbol{u} \in$ *Span S* which also violates our hypotheses and so is a contradiction. This completes the proof. □

What You Can Now Do

1. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of **n-vectors** determine if it is **linearly dependent** or **linearly independent**.

2. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of **n-vectors** is **linearly dependent**, find a **non-trivial dependence relation**.

3. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is **linearly dependent**, express one of the vectors as a **linear combination** of the remaining vectors.

Method (How do do it)

Method 2.4.1. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of **n-vectors** determine if it is **linearly dependent** or **linearly independent**.

Form the matrix A with the the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of **n-vectors** as its columns. Use **Gaussian elimination** to obtain an **echelon form** . If every **column** is a **pivot column** then the sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is **linearly independent**, otherwise it is **linearly dependent**.

Example 2.4.3. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ -3 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -4 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ -1 \\ -2 \\ 3 \end{pmatrix}$. Determine if the sequence $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**.

Form the matrix with these vectors as columns: $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \\ 1 & 1 & -2 \\ -3 & -4 & 3 \end{pmatrix}$. Using **Gaussian elimination** we obtain an **echelon form** :

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Each column is a **pivot column** and therefore the sequence S is **linearly independent**.

Example 2.4.4. Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ be as in **Example** (2.4.3) and $\boldsymbol{v}_4 = \begin{pmatrix} 2 \\ 2 \\ -1 \\ 0 \end{pmatrix}$. Determine if the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ **linearly independent**.

Now the matrix we must work with is 4×4 :

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

As before we apply **Gaussian elimination** to obtain an **echelon form** :

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

So the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly independent**.

Example 2.4.5. Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ be as in **Example** (2.4.3) and set $\boldsymbol{v}_4' = \begin{pmatrix} 2 \\ 2 \\ -1 \\ 1 \end{pmatrix}$.

Determine if $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4')$ is **linearly independent**.

Now the matrix we must consider is

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

Using **Gaussian elimination** we get an **echelon form** :

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

So in this case we conclude that the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4')$ is **linearly dependent**.

Method 2.4.2. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of **n-vectors** is **linearly dependent**, find a **non-trivial dependence relation**.

Assume using **Method** (2.4.1) we determine that the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is **linearly dependent**. To find a **non-trivial dependence relation** continue to use **Gaussian elimination** to obtain the **reduced echelon form** . Then use **Method** (2.3.3) to find a **nontrivial solution** to the **homogeneous linear system** which has **coefficient matrix** $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_k)$. Write out the homogeneous linear system corresponding to this matrix, identify the **leading variables** and **free variables** of the system. Set the free variables equal to parameters and express all the variables in terms of the parameters.

Example 2.4.6. We continue with **Example** (2.4.5). There we obtained the **echelon form**
$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 3 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. We apply **Gaussian elimination** to obtain the **reduced echelon form**

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

The **homogeneous linear system** with coefficient matrix $\begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ is

$$\begin{array}{ccccccc} x_1 & & & + & 2x_4 & = & 0 \\ & x_2 & & - & x_4 & = & 0 \\ & & x_3 & + & x_4 & = & 0 \end{array}$$

There are three **leading variables** (x_1, x_2, x_3) and a single **free variable** (x_4). Setting $x_4 = t$ we solve for each variable in terms of t and obtain the **general solution**:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2t \\ t \\ -t \\ t \end{pmatrix}.$$

Taking $t = 1$ we get the specific solution $\begin{pmatrix} -2 \\ 1 \\ -1 \\ 1 \end{pmatrix}$ which gives the **non-trivial dependence relation**

$$(-2)\boldsymbol{v}_1 + \boldsymbol{v}_2 + (-1)\boldsymbol{v}_3 + \boldsymbol{v}_4 = \mathbf{0}_4.$$

Example 2.4.7. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 5 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -6 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 2 \\ 3 \\ 1 \\ -4 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 3 \end{pmatrix}$.

Verify that $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent** and find a **non-trivial dependence relation**

We form the matrix with these vectors as columns, apply **Gaussian elimination** and obtain an **echelon form** :

$$A = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 2 & 1 & 3 & 0 \\ 3 & -1 & 1 & 1 \\ 5 & -6 & -4 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & -4 & -5 & 1 \\ 0 & -11 & -14 & 3 \end{pmatrix} \to$$

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

At this point we have confirmed that S is **linearly dependent** since the fourth column is not a **pivot column** . We continue using **Gaussian elimination** to obtain the **reduced echelon form**

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

The **homogeneous linear system** which has the above matrix as its **coefficient matrix** is

$$\begin{array}{rrrrrrr} x_1 & & & + & x_4 & = & 0 \\ & x_2 & & + & x_4 & = & 0 \\ & & x_3 & - & x_4 & = & 0 \end{array}$$

There are three **leading variables** (x_1, x_2, x_3) and a single **free variable** (x_4). Setting $x_4 = t$ we solve for each variable in terms of t and obtain the **general solution**:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -t \\ -t \\ t \\ t \end{pmatrix}$$

In particular, by taking $t = 1$ we get the specific **non-trivial dependence relation**:

$$(-1)\boldsymbol{v}_1 + (-1)\boldsymbol{v}_2 + \boldsymbol{v}_3 + \boldsymbol{v}_4 = \boldsymbol{0}_4.$$

Method 2.4.3. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is **linearly dependent**, express one of the vectors as a **linear combination** of the remaining vectors.

If one has a **non-trivial dependence relation** on the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \boldsymbol{0}$$

then for any coefficient which is not zero you can solve for the corresponding vector as a **linear combination** of the remaining vectors.

Example 2.4.8. Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ be as in **Example** (2.4.7). We have seen that

$$(-1)\boldsymbol{v}_1 + (-1)\boldsymbol{v}_2 + \boldsymbol{v}_3 + \boldsymbol{v}_4 = \boldsymbol{0}_4.$$

Since each coefficient is non-zero each vector of the sequence can be expressed as a **linear combination** of the others:

$$\boldsymbol{v}_1 = -\boldsymbol{v}_2 + \boldsymbol{v}_3 + \boldsymbol{v}_4$$

$$\boldsymbol{v}_2 = -\boldsymbol{v}_1 + \boldsymbol{v}_3 + \boldsymbol{v}_4$$

$$\boldsymbol{v}_3 = \boldsymbol{v}_1 + \boldsymbol{v}_2 - \boldsymbol{v}_4$$

$$\boldsymbol{v}_4 = \boldsymbol{v}_1 + \boldsymbol{v}_2 - \boldsymbol{v}_3.$$

Example 2.4.9. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -2 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ -1 \\ -4 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 3 \\ -5 \\ 1 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 0 \\ 1 \\ -3 \\ 2 \end{pmatrix}$.

Demonstrate that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent**. Find a **non-trivial dependence relation** and then express one of the vectors in terms of the vectors which precedes it.

We form the matrix $\begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 3 & 3 & 1 \\ -2 & -1 & -5 & -3 \\ -1 & -4 & 1 & 2 \end{pmatrix}$. Applying **Gaussian elimination** we obtain the **reduced echelon form** :

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

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

The **homogeneous linear system** which $\begin{pmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ as its **coefficient matrix** is

$$\begin{array}{ccccccccc} x_1 & & & + & 3x_3 & + & 2x_4 & = & 0 \\ & & x_2 & - & x_3 & - & x_4 & = & 0 \end{array}$$

There are two **leading variables** (x_1, x_2) and two **free variables** (x_3, x_4). Setting $x_3 = s, x_4 = t$ we solve for each variable in terms of s and t and obtain the **general solution**:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -3s - 2t \\ s + t \\ s \\ t \end{pmatrix}$$

Setting $s = 1, t = 0$ we get the **non-trivial dependence relation**

$$-3\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3 = \boldsymbol{0}_4.$$

We solve for $\boldsymbol{v}_3$ in terms of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$:

$$\boldsymbol{v}_3 = 3\boldsymbol{v}_1 - \boldsymbol{v}_2.$$

Exercises

In exercises 1 - 12 determine if the given sequence of **n-vectors** is **linearly independent**. If the exercise can be done without computation give an explanation for your assertion. See **Method** (2.4.1)

1. $\left(\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}2\\4\\-6\end{pmatrix}\right)$

2. $\left(\begin{pmatrix}1\\-2\\4\end{pmatrix}, \begin{pmatrix}3\\-5\\1\end{pmatrix}, \begin{pmatrix}2\\-3\\-2\end{pmatrix}\right)$

3. $\left(\begin{pmatrix}3\\-5\\1\end{pmatrix}, \begin{pmatrix}2\\-3\\-2\end{pmatrix}, \begin{pmatrix}1\\-2\\4\end{pmatrix}\right)$

4. $\left(\begin{pmatrix}1\\-2\\4\\1\end{pmatrix}, \begin{pmatrix}3\\-5\\1\\2\end{pmatrix}, \begin{pmatrix}2\\-3\\-2\\3\end{pmatrix}\right)$

5. $\left(\begin{pmatrix}1\\-2\\4\\1\end{pmatrix}, \begin{pmatrix}3\\-5\\1\\2\end{pmatrix}, \begin{pmatrix}2\\-3\\-2\\3\end{pmatrix}, \begin{pmatrix}0\\0\\1\\2\end{pmatrix}\right)$

6. $\left(\begin{pmatrix}1\\-2\\4\\1\end{pmatrix}, \begin{pmatrix}3\\-5\\1\\2\end{pmatrix}, \begin{pmatrix}0\\0\\1\\2\end{pmatrix}, \begin{pmatrix}2\\4\\5\\-4\end{pmatrix}\right)$

7. $\left(\begin{pmatrix}1\\-2\\4\\1\end{pmatrix}, \begin{pmatrix}3\\-5\\1\\2\end{pmatrix}, \begin{pmatrix}2\\4\\5\\-4\end{pmatrix}, \begin{pmatrix}1\\2\\3\\4\end{pmatrix}\right)$

8. $\left(\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}2\\3\\1\end{pmatrix}, \begin{pmatrix}3\\1\\2\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right)$

9. $\left(\begin{pmatrix}1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}4\\3\\2\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\end{pmatrix}\right)$

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

11. $\left(\begin{pmatrix}1\\4\\2\\0\\1\end{pmatrix}, \begin{pmatrix}3\\1\\5\\1\\0\end{pmatrix}, \begin{pmatrix}2\\-3\\3\\1\\-1\end{pmatrix}\right)$

12. $\left(\begin{pmatrix}1\\4\\2\\0\\0\end{pmatrix}, \begin{pmatrix}3\\1\\5\\0\\0\end{pmatrix}, \begin{pmatrix}2\\-2\\3\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\0\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\0\\1\\2\end{pmatrix}\right)$

In exercises 13-16 verify that the sequence of **n-vectors** is **linearly dependent** and find a **non-trivial dependence relation**.

See **Method** (2.4.1) and **Method** (2.4.2).

13. $\left(\begin{pmatrix}1\\2\\-3\end{pmatrix}, \begin{pmatrix}2\\-3\\1\end{pmatrix}, \begin{pmatrix}-3\\1\\2\end{pmatrix}\right)$

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

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

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

In exercises 17 - 18:

a) Verify that the sequence of **n-vectors** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**.
b) Demonstrate that sequence of **n-vectors** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent**.
c) Express $\boldsymbol{v}_4$ as a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

See **Method** (2.4.1), **Method** (2.4.2), and See **Method** (2.4.3).

17. $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 3 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 2 \\ 3 \\ 2 \\ 7 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

18. $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 2 \\ 2 \\ -3 \\ 1 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 3 \\ -2 \\ 1 \\ 2 \end{pmatrix}$

In exercises 19 - 24 answer true or false and give an explanation.

19. Any sequence of four vectors from $\mathbb{R}^5$ is **linearly independent**.

20. Any sequence of five vectors from $\mathbb{R}^5$ is **linearly dependent**. .

21. Some sequence of five vectors from $\mathbb{R}^5$ is **linearly dependent**.

22. Some sequence of five vectors from $\mathbb{R}^5$ is **linearly independent**.

23. Any sequence of six vectors from $\mathbb{R}^5$ is **linearly dependent**.

24. If $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ are in $\mathbb{R}^6$ and $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are linearly independent then $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ is linearly independent.

Challenge Exercises (Problems)

1. a) Show that the sequence consisting of the columns of the matrix

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

is **linearly independent**.

b) Assume now that the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**. Set

$$\boldsymbol{w}_1 = \boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 3\boldsymbol{v}_3, \boldsymbol{w}_2 = 2\boldsymbol{v}_1 + 3\boldsymbol{v}_2 + 4\boldsymbol{v}_3, \boldsymbol{w}_3 = -\boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 4\boldsymbol{v}_3.$$

Prove that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent**.

c) Show that the columns of the matrix $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ 3 & 4 & 1 \end{pmatrix}$ are **linearly dependent** and find a **non-trivial dependence relation**. See **Method** (2.4.1) and **Method** (2.4.2).

d) Set $\boldsymbol{u}_1 = \boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 3\boldsymbol{v}_3, \boldsymbol{u}_2 = 2\boldsymbol{v}_1 + 3\boldsymbol{v}_2 + 4\boldsymbol{v}_3, \boldsymbol{u}_3 = 1\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3$ and prove that the sequence $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ is **linearly dependent**.

2. Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ be **linearly independent** sequence of **3-vectors** . Prove that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a **spanning sequence** .

3. Assume the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ of **3-vectors** is a **spanning sequence** of $\mathbb{R}^3$. Prove that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**.

4.a) Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$ and set $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Demonstrate that S is **linearly independent** and is a **spanning sequence** of $\mathbb{R}^3$.

b) Let $V = \{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4 : x_1 + x_2 + x_3 - x_4 = 0\}$. Prove that V is a **subspace** of $\mathbb{R}^4$.

c) Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 3 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 2 \\ 1 \\ -1 \\ 2 \end{pmatrix}$ and set $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$. Prove that $\mathcal{B}$ is **linearly independent** and **spans** V. (Hints: (1) If $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \in V$ and $x_1 = y_1, x_2 = y_2, x_3 = y_3$ then prove that $x_4 = y_4$ and $\boldsymbol{x} = \boldsymbol{y}$. (2) Make use of part a).

5. Let $V = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : x + y - 2z = 0\}$.

a) Prove that V is a **subspace** of $\mathbb{R}^3$.

b) Assume that $\boldsymbol{v}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}$ are in V and $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is **linearly independent**.

Set $\boldsymbol{w}_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$. Prove $(\boldsymbol{w}_1, \boldsymbol{w}_2)$ is **linearly independent** and a **spanning sequence** of $\mathbb{R}^2$.

c) Show that any sequence S of three vectors in V must be **linearly dependent**. (See hints to 4c).

d) Prove if T is a sequence of length two from V and T is **linearly independent** then T **spans** V.

Quiz Solutions

1. The **reduced echelon form** of A is $\begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. The **pivot positions** are (1,1), (2,2) and (3,4). The **pivot column** are 1, 2 and 4.

Not right, **Method** (1.2.3).

2. The sequence of columns of A is not a **spanning sequence** of $\mathbb{R}^4$ since not every row of the **reduced echelon form** of A has a **pivot position** (the fourth row does not have a pivot position).

Not right, see **Method** (2.3.1).

3. The number of **solutions** of the **homogeneous linear system** with A as its matrix is d) infinite.

Not right, see **Theorem** (1.2.4).

4. Assume that A is the **matrix** of a linear system. The number of **solutions** of the **inhomogeneous linear system** with A as its matrix is a) zero, the system is **inconsistent**.

Not right, see **Theorem** (1.2.1).

5. The solution set is $Span\left(\begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix}\right)$.

Not right, see **Method** (1.2.4) and **Method** (2.3.3).

6. The number of **solutions** of the **inhomogeneous linear system** with **augmented matrix**

$\begin{pmatrix} 1 & 2 & 1 & 3 & | & 1 \\ 1 & 3 & 0 & 4 & | & 0 \\ 1 & 1 & 2 & 2 & | & 2 \\ 1 & -1 & 4 & 1 & | & 6 \end{pmatrix}$ is d) infinite.

Not right, see **Theorem** (1.2.3) and **Theorem** (2.2.2).

2.5. Subspaces and Bases of $\mathbb{R}^n$

In this section we study subspaces of $\mathbb{R}^n$ and introduce the notion of a basis. We show every subspace of $\mathbb{R}^n$ has a basis and the number of elements of a basis (its length) is unique. We also introduce the concept of a coordinate vector with respect to a basis.

Am I Ready for This Material

In order to follow this section and master its ideas and methods you need to be familiar with many previously defined concepts. **Those that are used extensively in this section are**:

linear system

solution of a linear system

solution set of a linear system

consistent linear system

inconsistent linear system

equivalent linear systems

echelon form of a linear system

leading variable

free variable

matrix

augmented matrix of a linear system

row equivalence of matrices

matrix in row echelon form

matrix in reduced row echelon form

echelon form of a matrix

reduced echelon from of a matrix

pivot positions of a matrix

pivot columns of a matrix

an $n-$vector

equality of $n-$vectors

$n-$space, $\mathbb{R}^n$

adddition of $n-$vectors

scalar multiplication of an $n-$vector

negative of a vector

the zero vector

the standard basis for $\mathbb{R}^n$

linear combination of vectors

span of a sequence $(v_1, v_2, \dots, v_k)$ from $\mathbb{R}^n$

spanning sequence of $\mathbb{R}^n$

subspace of $\mathbb{R}^n$

spanning sequence of a subspace V of $\mathbb{R}^n$

dependence relation on a sequence of n-vectors

trivial and non-trivial dependence relation on a sequence of n-vectors

linearly dependent sequence of n-vectors

linearly independent sequence of n-vectors

You will also need to be able to execute several procedures previously described. **The following are algorithms that we will use in this section**:

Gaussian elimination

procedure for solving a linear system

Method (2.3.2): Determine if a sequence $(v_1, \dots, v_k)$ of **n-vectors** is a **spanning sequence** of $\mathbb{R}^n$

Method (2.4.1): Determine if a sequence $(v_1, \dots, v_k)$ is **linearly independent**.

Quiz

Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 0 \\ -5 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 2 \\ 1 \\ 3 \\ -6 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ -1 \\ 2 \\ -2 \end{pmatrix}, \boldsymbol{v}_5 = \begin{pmatrix} 1 \\ -1 \\ 4 \\ -3 \end{pmatrix}$.

1. Verify that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**.

2. Demonstrate that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent** and find a **non-trival dependence relation**.

3. Express $\boldsymbol{v}_4$ as a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

4. Demonstrate that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_5)$ is a **spanning sequence** of $\mathbb{R}^4$.

5. Express each of the **standard basis vectors** , $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4$, as a linear combination of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_5)$.

6. Demonstrate that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_5)$ is **linearly independent**.

Quiz Solutions

New Concepts

In this section we introduce three entirely new concepts. These are:

For a **subspace** W of $\mathbb{R}^n$ the sequence $\mathcal{B} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is a **basis** of W.

For a **subspace** W of $\mathbb{R}^n$ the **dimension** of W.

For a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ of a **subspace** W of $\mathbb{R}^n$ and an **n-vector** $\boldsymbol{w}$ of W, the **coordinate vector** of $\boldsymbol{w}$ with respect to $\mathcal{B}$.

Theory (Why it Works)

In Chapter 5 we will introduce a new, abstract concept, that of a **vector space**. The first examples will be $\mathbb{R}^n$ and its **subspaces** . In order to get a feel for the sorts of things we will be proving we first study in some detail $\mathbb{R}^n$ and its **subspaces** . Recall that we previously showed that addition and scalar multiplication in $\mathbb{R}^n$ satisify the following properties:

For **n-vectors** $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ and scalars a, b the following hold

1. $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$.

2. $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$.

3. $\boldsymbol{u}+\mathbf{0}=\boldsymbol{u}$.

4. $\boldsymbol{u}+(-\boldsymbol{u})=\mathbf{0}$.

5. $(a+b)\boldsymbol{u}=a\boldsymbol{u}+b\boldsymbol{u}$.

6. $a(\boldsymbol{u}+\boldsymbol{v})=a\boldsymbol{u}+a\boldsymbol{v}$.

7. $(ab)\boldsymbol{u}=a(b\boldsymbol{u})$.

8. $1\boldsymbol{u}=\boldsymbol{u}$.

9. $0\boldsymbol{u}=\mathbf{0}$.

We are interested in those nonempty subsets of $\mathbb{R}^n$ which satisfy these same properties, with the **addition** and **scalar multiplication** "inherited" from $\mathbb{R}^n$. We have briefly encountered these before - they are the **subspaces** $\mathbb{R}^n$.

Remark 2.9. If V is a **subspace** of $\mathbb{R}^n$ then $\mathbf{0}_n \in V$. To see this, note that by assumption 1) V is nonempty and therefore there is some vector $\boldsymbol{v} \in V$. Then by 2) the **scalar product** of $\boldsymbol{v}$ by any scalar $c, c\boldsymbol{v}$ is in V. In particular, we can take $c=0$ and then $\mathbf{0}_n=0\boldsymbol{v} \in V$.

Example 2.5.1. The subset $\{\mathbf{0}_n\}$ of $\mathbb{R}^n$ is a **subspace** .

Definition 2.29. The **subspace** $\{\mathbf{0}_n\}$ of $\mathbb{R}^n$ is referred to as the ***zero subspace*** of $\mathbb{R}^n$. Any other **subspace** of $\mathbb{R}^n$ is said to be ***non-zero***.

Example 2.5.2. Set $\boldsymbol{v}(x,y)=\begin{pmatrix} x+2y \\ 2x-y \\ x+y \end{pmatrix}$. Determine if $V=\{\boldsymbol{v}(x,y): x,y \in \mathbb{R}\}$ is a **subspace** of $\mathbb{R}^3$.

Let

$$\boldsymbol{u}=\boldsymbol{v}(x_1,y_1)=\begin{pmatrix} x_1+2y_1 \\ 2x_1-y_1 \\ x_1+y_1 \end{pmatrix}, \boldsymbol{w}=\boldsymbol{v}(x_2,y_2)=\begin{pmatrix} x_2+2y_2 \\ 2x_2-y_2 \\ x_2+y_2 \end{pmatrix}.$$

Then

$$\boldsymbol{u}+\boldsymbol{v}=\begin{pmatrix} x_1+2y_1 \\ 2x_1-y_1 \\ x_1+y_1 \end{pmatrix}+\begin{pmatrix} x_2+2y_2 \\ 2x_2-y_2 \\ x_2+y_2 \end{pmatrix}=\begin{pmatrix} (x_1+2y_1)+(x_2+2y_2) \\ (2x_1-y_1)+(2x_2-y_2) \\ (x_1+y_1)+(x_2+y_2) \end{pmatrix}=$$

$$\begin{pmatrix} (x_1+x_2)+2(y_1+y_2) \\ 2(x_1+x_2)-(y_1+y_2) \\ (x_1+x_2)+(y_1+y_2) \end{pmatrix} = \boldsymbol{v}(x_1+x_2, y_1+y_2)$$

which belongs to V.

Let $\boldsymbol{z} = \boldsymbol{v}(x,y) = \begin{pmatrix} x+2y \\ 2x-y \\ x+y \end{pmatrix}$ and c be a scalar. Then

$c\boldsymbol{w} = c\begin{pmatrix} x+2y \\ 2x-y \\ x+y \end{pmatrix} = \begin{pmatrix} c(x+2y) \\ c(2x-y) \\ c(x+y) \end{pmatrix} = \begin{pmatrix} (cx)+2(cy) \\ 2(cx)-(cy) \\ (cx)+(cy) \end{pmatrix} = \boldsymbol{v}(cx, cy) \in V.$

Example 2.5.3. Is $V = \{\begin{pmatrix} x^2 \\ 2x \\ x \end{pmatrix} : x \in \mathbb{R}\}$ a subspace of $\mathbb{R}^3$?

The vector $\boldsymbol{v} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ is in V. We claim that the $2\boldsymbol{v} = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix}$ is not in V. For suppose there is an x such that $\begin{pmatrix} x^2 \\ 2x \\ x \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix}$. Comparing the last component we see that we must have $x = 2$. But then the first component is $2^2 = 4 \neq 2$ and so the scalar multiple $2\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ is not in V. Thus, V is not a **subspace** .

Example 2.5.4. Let $V = \{\begin{pmatrix} 2x-2y+1 \\ x-y \\ x+y \end{pmatrix} : x, y \in \mathbb{R}\}$. Is V a **subspace** of $\mathbb{R}^3$?

If V is a **subspace** then V contains the **zero vector** , $\mathbf{0}_3$. However, if $\begin{pmatrix} 2x-2y+1 \\ x-y \\ x+y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ then $\begin{pmatrix} x \\ y \end{pmatrix}$ is a **solution** to the **linear system**

$$\begin{array}{rcrcr} 2x & - & 2y & = & -1 \\ x & - & y & = & 0 \\ x & + & y & = & 0 \end{array}$$

However, this system is **inconsistent** (check this) and so the **zero vector** does not belong to V and V is not a **subspace** .

Example 2.5.5. We have previously seen that there are lots of **subspaces** of $\mathbb{R}^n$; specifically, we proved in **Theorem** (2.3.5) if $S = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is any sequence of **n-vectors** then $Span\ S$ is a subspace.

The next theorem describes a general way to obtain **subspaces** . The meaning of this theorem will become clearer in Chapter 3 when we define ***matrix multiplication*** and the ***null space*** of a matrix.

Theorem 2.5.1. *Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ be a sequence of **n-vectors** . Let V consist of those **m-vectors** $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$ such that $x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \dots x_m\boldsymbol{v}_m = \boldsymbol{0}_n$. Then V is a **subspace** of $\mathbb{R}^m$.*

Proof. For the purpose of exposition we do the proof in the case that $m = 4$; the structure of the proof for general m is exactly the same.

First of all V is non-empty since the **zero vector** , $\boldsymbol{0}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ is in V :

$$0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 0\boldsymbol{v}_3 + 0\boldsymbol{v}_4 = \boldsymbol{0}_n.$$

We have to show that V is **closed under addition and scalar multiplication**.

Closed under addition: Suppose that $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \in V$. This means that

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

$$y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 + y_3\boldsymbol{v}_3 + y_4\boldsymbol{v}_4 = \boldsymbol{0}_n.$$

We need to show that $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ x_4 + y_4 \end{pmatrix} \in V$ which means we need to show that

$$(x_1 + y_1)\boldsymbol{v}_1 + (x_2 + y_2)\boldsymbol{v}_2 + (x_3 + y_3)\boldsymbol{v}_3 + (x_4 + y_4)\boldsymbol{v}_4 = \boldsymbol{0}_n.$$

However, by the distributive property

$$(x_1+y_1)\boldsymbol{v}_1+(x_2+y_2)\boldsymbol{v}_2+(x_3+y_3)\boldsymbol{v}_3+(x_4+y_4)\boldsymbol{v}_4 =$$
$$(x_1\boldsymbol{v}_1+y_1\boldsymbol{v}_1)+(x_2\boldsymbol{v}_2+y_2\boldsymbol{v}_2)+(x_3\boldsymbol{v}_3+y_3\boldsymbol{v}_3)+(x_4\boldsymbol{v}_4+y_4\boldsymbol{v}_4)$$

In turn, using commutativity and associativity of addition we get

$$(x_1\boldsymbol{v}_1+y_1\boldsymbol{v}_1)+(x_2\boldsymbol{v}_2+y_2\boldsymbol{v}_2)+(x_3\boldsymbol{v}_3+y_3\boldsymbol{v}_3)+(x_4\boldsymbol{v}_4+y_4\boldsymbol{v}_4) =$$

$$(x_1\boldsymbol{v}_1+x_2\boldsymbol{v}_2+x_3\boldsymbol{v}_3+x_4\boldsymbol{v}_4)+(y_1\boldsymbol{v}_1+y_2\boldsymbol{v}_2+y_3\boldsymbol{v}_3+y_4\boldsymbol{v}_4) = \mathbf{0}_n+\mathbf{0}_n=\mathbf{0}_n$$

as required.

Closed under scalar multiplication: Assume $\begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4\end{pmatrix} \in V$ and c is a scalar. We need to show that $c\begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4\end{pmatrix} = \begin{pmatrix} cx_1\\ cx_2\\ cx_3\\ cx_4\end{pmatrix} \in V$.

We therefore have to show that

$$(cx_1)\boldsymbol{v}_1+(cx_2)\boldsymbol{v}_2+(cx_3)\boldsymbol{v}_3+(cx_4)\boldsymbol{v}_4=\mathbf{0}_n.$$

However, using the general property that for scalars a, b and a vector $\boldsymbol{u}$ that $(ab)\boldsymbol{u} = a(b\boldsymbol{u})$ we get

$$(cx_1)\boldsymbol{v}_1+(cx_2)\boldsymbol{v}_2+(cx_3)\boldsymbol{v}_3+(cx_4)\boldsymbol{v}_4 = c(x_1\boldsymbol{v}_1)+c(x_2\boldsymbol{v}_2)+c(x_3\boldsymbol{v}_3)+c(x_4\boldsymbol{v}_4).$$

Using the distributive property we get

$$c(x_1\boldsymbol{v}_1)+c(x_2\boldsymbol{v}_2)+c(x_3\boldsymbol{v}_3)+c(x_4\boldsymbol{v}_4) = c[x_1\boldsymbol{v}_1+x_2\boldsymbol{v}_2+x_3\boldsymbol{v}_3+x_4\boldsymbol{v}_4] = c\mathbf{0}_n=\mathbf{0}_n$$

since $\begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4\end{pmatrix} \in V$ implies that $x_1\boldsymbol{v}_1+x_2\boldsymbol{v}_2+x_3\boldsymbol{v}_3+x_4\boldsymbol{v}_4=\mathbf{0}_n$. This completes the proof. $\square$

Later in this section we show that every **subspace** V of $\mathbb{R}^n$ can be expressed as the **span** of a **linearly independent** sequence of vectors. Such sequences are extremely important and the subject of the next definition:

Definition 2.30. Let V be a nonzero **subspace** of $\mathbb{R}^n$. A sequence $\mathcal{B}$ of V is a **basis** if

1. $Span\ \mathcal{B} = V$, that is, the $\mathcal{B}$ **spans** V; and

2. $\mathcal{B}$ is **linearly independent**.

Example 2.5.6. The sequence of **standard basis vectors** in $\mathbb{R}^n$, $(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)$, is a basis of $\mathbb{R}^n$.

In our next theorem we show that every **subspace** W of $\mathbb{R}^n$ has a **basis** . One consequence of this is that every **subspace** has the form $Span\ (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ for some sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ from W. We will then show that any two bases of a subspace W of $\mathbb{R}^n$ have the same length.

Theorem 2.5.2. *Let W be a* **non-zero subspace** *of $\mathbb{R}^n$. Then W has a* **basis** *(with at most n vectors).*

Proof. Since W is a **non-zero subspace** of $\mathbb{R}^n$, in particular, there exists a vector $\boldsymbol{w} \neq \boldsymbol{0}_n$ in W. Then the sequence $(\boldsymbol{w})$ is **linearly independent** by **Theorem** (2.4.5). Thus, we know there exists **linearly independent** sequences of vectors in W.

Let $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ be a **linearly independent**sequence of vectors from W with k chosen as large as possible. Since a **linearly independent**sequence in W is linearly independent in $\mathbb{R}^n$ we know by **Theorem** (2.4.2) that $k \leq n$. We claim that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ is a **spanning sequence** of W.

Suppose not and assume that $\boldsymbol{w}$ is in W but $\boldsymbol{w}$ is not in $Span(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k)$. Then by **Theorem** (2.4.11) the sequence $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k, \boldsymbol{w})$ is **linearly independent** and contained in W, which contradicts the choice of k (it is a linearly independent sequence with length greater than $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k)$). Therefore, no such $\boldsymbol{w}$ exists and $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ is a **spanning sequence** of W. Since $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k)$ is **linearly independent** it follows that is a **basis** $\square$

Example 2.5.7. Determine if the sequence of columns of the matrix $\begin{pmatrix} 1 & 4 & 0 \\ -2 & -5 & 3 \\ 1 & 7 & 6 \end{pmatrix}$ is a basis for $\mathbb{R}^3$.

By **Theorem** (2.4.4) ("half is good enough theorem") the sequence of the columns of an $n \times n$ matrix is a **basis** of $\mathbb{R}^n$ if only if every column is a **pivot column** . Thus, we need to apply **Gaussian elimination** to this matrix in order to obtain its **reduced echelon form** .

$$\to \begin{pmatrix} 1 & 4 & 0 \\ 0 & 3 & 3 \\ 0 & 3 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus, the sequence of columns $\left(\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ -5 \\ 7 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 6 \end{pmatrix} \right)$ is a **basis** for $\mathbb{R}^3$.

Example 2.5.8. Let

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \boldsymbol{v}_2 \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 4 \\ -1 \\ -3 \end{pmatrix}.$$

Is the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ a **basis** of $\mathbb{R}^3$.

We use the same procedure: Form the matrix with these vectors as its columns and use **Gaussian elimination** to determine if the **reduced echelon form** is the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. We do the calculation:

$$\begin{pmatrix} 1 & 1 & 4 \\ 1 & -3 & -1 \\ -2 & 2 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 4 \\ 0 & -4 & -5 \\ 0 & 4 & 5 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & \frac{5}{4} \\ 0 & 0 & 0 \end{pmatrix}$$

Therefore the sequence is not a **basis** .

Example 2.5.9. The set $S = \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} : x_1 + x_2 + x_3 + x_4 = 0\}$ is a **subspace** of $\mathbb{R}^4$. Show that the sequence of vectors $\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \right)$ is a basis of S.

Each of the vectors belongs to S since $1+(-1)=0$. The sequence is also **linearly independent** since the **reduced echelon form** of $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ and every column is a **pivot column** .

Finally, suppose $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in S$. Then $x_1+x_2+x_3+x_4=0$ from which it follows that $x_4=-x_1-x_2-x_3$. Then

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

which shows that the sequence of vectors $\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \right)$ is a **spanning sequence** of S.

Example 2.5.10. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$. Find a **basis** for the **subspace** V of $\mathbb{R}^4$ consisting of those vectors $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ such that $x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + x_3\boldsymbol{v}_3 + x_4\boldsymbol{v}_4 = \mathbf{0}_3$.

$$x_1 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} x_1+2x_2+x_3+x_4 \\ 2x_1+3x_2+x_3+3x_4 \\ x_1+x_2+2x_4 \end{pmatrix}$$

By setting this equal to the **zero vector** , $\mathbf{0}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ we obtain a **homogeneous linear**

system of three equations in the four variables x_1, x_2, x_3, x_4. The **coefficient matrix** of the system is the matrix which has $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ as its columns: $\begin{pmatrix} 1 & 2 & 1 & 1 \\ 2 & 3 & 1 & 3 \\ 1 & 1 & 0 & 2 \end{pmatrix}$.

The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -1 & 3 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. The corresponding homogeneous linear system is:

$$\begin{array}{ccccccccc} x_1 & & & - & x_3 & + & 3x_4 & = & 0 \\ & & x_2 & + & x_3 & - & x_4 & = & 0 \end{array}$$

This system has two **leading variable** (x_1, x_2) and two free variables **free variable** (x_3, x_4). We set $x_3 = s, x_4 = t$ and solve for all the variables in terms of the parameters s and t :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} s - 3t \\ -s + t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Consequently, $V = Span\left(\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right)$. Since $\left(\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right)$ is **linearly independent** (the vectors are not multiples of each other) it follows that this sequence is a **basis** for V.

Example 2.5.11. Let $V = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ -6 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \\ -4 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}\right)$. Determine if the sequence $\mathcal{B} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$ is a **basis** of V.

We have to do three things to determine this:

1) Show that each of the vectors of $\mathcal{B}$ belongs to V;
2) Show that $\mathcal{B}$ is **linearly independent**; and

3) Show that $\mathcal{B}$ is a **spanning sequence** of V.

To do 1) we have to determine if each of the vectors $\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$ is a **linear combination** of $\begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ -6 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \\ -4 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$ and we can do this all at once as follows: form the matrix whose first four columns are the vectors $\begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 2 \\ 3 \\ -6 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \\ -4 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$ augmented by the three vectors $\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$. Use **Gaussian elimination** to obtain an **echelon form** . If any of the three augmented columns is a **pivot column** then the corresponding vector is not in V and we can conclude that the sequence is not a **basis** . In the contrary case each of the vectors of the sequence belongs to V.

The required matrix is $\begin{pmatrix} 1 & 1 & 2 & 1 & | & 1 & 0 & 0 \\ 1 & 2 & 1 & -1 & | & -1 & 1 & 0 \\ 1 & 3 & 1 & 1 & | & 0 & -1 & 1 \\ -3 & -6 & -4 & -1 & | & 0 & 0 & -1 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 0 & -9 & | & -6 & 8 & -3 \\ 0 & 1 & 0 & 2 & | & 1 & -2 & 1 \\ 0 & 0 & 1 & 4 & | & 3 & -3 & 1 \\ 0 & 0 & 0 & 0 & | & 0 & 0 & 0 \end{pmatrix}$. Thus, we conclude that each of the three vectors of the sequence belong to V.

To determine if $\mathcal{B}$ is **linearly independent** we make a matrix with the three vectors as columns and use **Gaussian elimination** to find an **echelon form** . If each of the columns is a **pivot column** then $\mathcal{B}$ is independent, otherwise the sequence is **linearly dependent** .

To demonstrate 3) we must do something similar to 1):

We form the matrix whose first three columns are the vectors of the sequence $\left(\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}\right)$ augmented by the four vectors spanning V. If any of the four augmented columns is a **pivot column** then the corresponding vector is not in $Span\ \mathcal{B}$. In this case we would conclude that $\mathcal{B}$ does not span V and it not a basis. On the other hand, if no column on the right is a pivot column then $\mathcal{B}$ does span V.

We can actually do 2) and 3) all at once: the matrix for 3) is

$$\begin{pmatrix} 1 & 0 & 0 & | & 1 & 1 & 2 & 1 \\ -1 & 1 & 0 & | & 1 & 2 & 1 & -1 \\ 0 & -1 & 1 & | & 1 & 3 & 1 & 1 \\ 0 & 0 & -1 & | & -3 & -6 & -4 & -1 \end{pmatrix}.$$

Using **Gaussian elimination** we obtain the **reduced echelon form** of this matrix. If each of the first three columns is a **pivot column** then $\mathcal{B}$ is **linearly independent** and if neither of the last four columns is a pivot then $\mathcal{B}$ **spans** V and a **basis** of V. If either of these conditions does not hold then $\mathcal{B}$ is not a basis.

The **reduced echelon form** of the matrix is $\begin{pmatrix} 1 & 0 & 0 & | & 1 & 1 & 2 & 1 \\ 0 & 1 & 0 & | & 2 & 3 & 3 & 0 \\ 0 & 0 & 1 & | & 3 & 6 & 4 & 1 \\ 0 & 0 & 0 & | & 0 & 0 & 0 & 0 \end{pmatrix}$ and therefore $\mathcal{B}$ is a **basis** of V.

Theorem (2.4.4) (the “half is good enough theorem”), which was proved in the last section, implies that any basis of $\mathbb{R}^n$ has exactly n vectors. What about a **subspace** V of $\mathbb{R}^n$, do all subspaces of V has the same number of vectors. Toward showing that all bases of a subspace of $\mathbb{R}^n$ all have the same length we prove the following lemma which will be referred to as the “exchange theorem for $\mathbb{R}^n$”.

Lemma 2.5.3. *Let W be a **subspace** of $\mathbb{R}^n$ and assume $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_l)$ is a **spanning sequence** of W. If $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is a **linearly independent** sequence from W then $k \leq l$. Alternatively, if $k > l$ then $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is **linearly dependent**.*

Proof. Rather than prove the general case we prove it in the case that $l = 3$ which will have all the elements of a general proof. So, we need to prove if $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is a sequence of length four from W then it is **linearly dependent**.

Since $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is a **spanning sequence** for W we can write each $\boldsymbol{v}_j$ as a **linear combination** of $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$. Thus, assume

$$\boldsymbol{v}_1 = a_{11}\boldsymbol{w}_1 + a_{21}\boldsymbol{w}_2 + a_{31}\boldsymbol{w}_3$$

$$\boldsymbol{v}_2 = a_{12}\boldsymbol{w}_1 + a_{22}\boldsymbol{w}_2 + a_{32}\boldsymbol{w}_3$$

$$\boldsymbol{v}_3 = a_{13}\boldsymbol{w}_1 + a_{23}\boldsymbol{w}_2 + a_{33}\boldsymbol{w}_3$$

$$\boldsymbol{v}_4 = a_{14}\boldsymbol{w}_1 + a_{24}\boldsymbol{w}_2 + a_{34}\boldsymbol{w}_3$$

Consider the sequence $\left(\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix}, \begin{pmatrix} a_{14} \\ a_{24} \\ a_{34} \end{pmatrix}\right)$ from $\mathbb{R}^3$. By **Theorem** (2.4.3) this sequence is **linearly dependent** . Consequently, there is a non-zero **4-vector** , $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$, such that

$$c_1 \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + c_2 \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} + c_3 \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} + c_4 \begin{pmatrix} a_{14} \\ a_{24} \\ a_{34} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

After performing the vector operations we conclude that

$$\begin{pmatrix} c_1a_{11} + c_2a_{12} + c_3a_{13} + c_4a_{14} \\ c_1a_{21} + c_2a_{22} + c_3a_{23} + c_4a_{24} \\ c_1a_{31} + c_2a_{32} + c_3a_{33} + c_4a_{34} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Now consider $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4$. Writing this in terms of $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ we obtain

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 =$$

$$c_1(a_{11}\boldsymbol{w}_1 + a_{21}\boldsymbol{w}_2 + a_{31}\boldsymbol{w}_3) \ + \ c_2(a_{12}\boldsymbol{w}_1 + a_{22}\boldsymbol{w}_2 + a_{32}\boldsymbol{w}_3) +$$

$$c_3(a_{13}\boldsymbol{w}_1 + a_{23}\boldsymbol{w}_2 + a_{33}\boldsymbol{w}_3) \ + \ c_4(a_{14}\boldsymbol{w}_1 + a_{24}\boldsymbol{w}_2 + a_{34}\boldsymbol{w}_3).$$

After distributing and commuting we can write this as

$$(c_1a_{11} + c_2a_{12} + c_3a_{13} + c_4a_{14})\boldsymbol{w}_1 \ + \ (c_1a_{21} + c_2a_{22} + c_3a_{23} + c_4a_{24})\boldsymbol{w}_2 +$$

$$(c_1a_{31} + c_2a_{32} + c_3a_{33} + c_4a_{34})\boldsymbol{w}_3 =$$

$$0\boldsymbol{w}_1 + 0\boldsymbol{w}_2 + 0\boldsymbol{w}_3 = \boldsymbol{0}_n.$$

Thus, we have explicitly demonstrated that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent** . $\square$

Theorem 2.5.4. *Let W be a* **subspace** *of* $\mathbb{R}^n$ *. Then every* **basis** *of W has the same number of elements.*

Proof. Let $\mathcal{B} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_k)$ and $\mathcal{B}' = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_l)$ be two bases of W. Since $\mathcal{B}$ is a basis of W it is a **spanning sequence**. Since $\mathcal{B}'$ is a basis of W it is **linearly**

independent. By **Lemma** (2.5.3) it follows that $l \leq k$. On the other hand, since $\mathcal{B}'$ is a basis of W it is **spanning sequence**, and since $\mathcal{B}$ is a basis of W it is **linearly independent**. Again by **Lemma** (2.5.3) we conclude that $k \leq l$. Thus, $k = l$. $\square$

Definition 2.31. Let V be a **subspace** of $\mathbb{R}^n$. Then the number of vectors in a **basis** of V is called the ***dimension*** of V.

Example 2.5.12. The dimension of $\mathbb{R}^n$ is n since the **standard basis** of $\mathbb{R}^n$ is a **basis** .

Example 2.5.13. By **Example** (2.5.9) the **subspace**

$$S = \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} : x_1 + x_2 + x_3 + x_4 = 0 \}$$

of $\mathbb{R}^4$ has dimension three since the sequence $\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \right)$ is a **basis** of S.

Example 2.5.14. The **subspace** V of **Example** (2.5.10) has dimension two.

Example 2.5.15. The subspace in **Example** (2.5.11) has dimension three.

One reason that the concept of a **basis** for a **subspace** V of $\mathbb{R}^n$ is so important is that every vector $\boldsymbol{v}$ of the subspace V can be written in a unique way as a **linear combination** of the basis vectors. We demonstrate this in the next theorem. This will provide a convenient way of expressing the vectors in the subspace.

Theorem 2.5.5. *Let V be a* **subspace** *of $\mathbb{R}^n$. A sequence $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is a* **basis** *of V if and only if every vector in V is uniquely a* **linear combination** *of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.*

Proof. Assume every vector in V is uniquely a **linear combination** of the vectors in $\mathcal{B}$. So, by assumption, $\mathcal{B}$ is a **spanning sequence** of V and so it remains to show that

$\mathcal{B}$ is **linearly independent**. Also, since the **zero vector** , $\mathbf{0}_n$, is in V by the uniqueness assumption there is only one **linear combination** of the form

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}_n. \tag{2.16}$$

However, the **trivial dependence relation** satisfies (2.16) so by uniqueness the only dependence relation on $\mathcal{B}$ is this one and, consequently, $\mathcal{B}$ is **linearly independent** and hence a **basis** .

Conversely, assume that $\mathcal{B}$ is a basis. Then $\mathcal{B}$ is a **spanning sequence** of V which means that that every vector in V is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$. So we must show uniqueness. Toward that end, suppose that

$a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_k\boldsymbol{v}_k$. We need to show that $a_1 = b_1, a_2 = b_2, \ldots, a_k = b_k$.

Since we have the above equality we can subtract $b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_k\boldsymbol{v}_k$ from both sides to get

$$(a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k) - (b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_k\boldsymbol{v}_k) = \mathbf{0}.$$

The left hand side can be rearranged:

$$(a_1 - b_1)\boldsymbol{v}_1 + (a_2 - b_2)\boldsymbol{v}_2 + \cdots + (a_k - b_k)\boldsymbol{v}_k = \mathbf{0}_n.$$

This is a **dependence relation** which must be trivial since $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is **linearly independent**. This implies that

$$a_1 - b_1 = a_2 - b_2 = \ldots a_k - b_k = 0.$$

Since for each $i, a_i - b_i = 0$ we conclude that

$$a_1 = b_1, a_2 = b_2, \ldots, a_k = b_k$$

from which the uniqueness now follows.

$\square$

Definition 2.32. Let V be a **subspace** of $\mathbb{R}^n$, $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be a **basis** , and $\boldsymbol{v}$ a vector in V. Assume $\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots c_k\boldsymbol{v}_k$. The unique scalars $c_1, c_2, \ldots, c_k$ are called the ***coordinates of $\boldsymbol{v}$ with respect to $\mathcal{B}$***. The vector with components $c_1, c_2, \ldots, c_k$ is called the ***coordinate vector of $\boldsymbol{v}$ with respect to $\mathcal{B}$*** and is denoted by

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

Example 2.5.16. In **Example** (2.5.11) we demonstrated that the sequence of vectors $\mathcal{B} = \left(\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \right)$ is **linearly independent**. Notice that each of these belongs to the subspace S of **Example** (2.5.9) which was shown to have dimension 3. We claim that $\mathcal{B}$ must **span** S. If not, then there is some vector $\boldsymbol{u} \in S, \boldsymbol{u} \notin Span\mathcal{B}$. But then extending $\mathcal{B}$ by $\boldsymbol{u}$ makes a **linearly independent** sequence with four vectors in S, which is not possible by the proof of the **Theorem** (2.5.4). Therefore $\mathcal{B}$ is, indeed, a basis.

We now determine the **coordinate vector** $\begin{pmatrix} 3 \\ 2 \\ 1 \\ -6 \end{pmatrix}$ with respect to $\mathcal{B}$.

Make the matrix with first three columns the vectors of the **basis** $\mathcal{B}$ and then augment by the vector $\begin{pmatrix} 3 \\ 2 \\ 1 \\ -6 \end{pmatrix}$. Apply **Gaussian elimination** to determine the **reduced echelon form** of this matrix. This will yield a unique solution to the **linear system** which has this matrix as its **augmented matrix** and this solution is the **coordinate vector** .

The matrix of this system is $\begin{pmatrix} 1 & 0 & 0 & | & 3 \\ -1 & 1 & 0 & | & 2 \\ 0 & -1 & 1 & | & 1 \\ 0 & 0 & -1 & | & -6 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 0 & | & 3 \\ 0 & 1 & 0 & | & 5 \\ 0 & 0 & 1 & | & 6 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$ and therefore the **coordinate vector** of $\begin{pmatrix} 3 \\ 2 \\ 1 \\ -6 \end{pmatrix}$ with respect to $\mathcal{B}$ is $\begin{pmatrix} 3 \\ 5 \\ 6 \end{pmatrix}$.

What You Can Now Do

1. Determine if a given subset of $\mathbb{R}^n$ is a **subspace**.

rmine if $\boldsymbol{u} \in V$.

3. Given a **subspace** V of $\mathbb{R}^n$ and a sequence $\mathcal{B} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ of V determine if $\mathcal{B}$ is a basis of V.

4. Given a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ for a **subspace** V of $\mathbb{R}^n$ and an **n-vector** , $\boldsymbol{v} \in V$ determine the **coordinate vector** , $[\boldsymbol{v}]_{\mathcal{B}}$, of $\boldsymbol{v}$ with respect to the basis $\mathcal{B}$.

Method (How to do it)

Method 2.5.1. Determine if a given subset V of $\mathbb{R}^n$ is a **subspace** .

Typically such a set is given explicitly, meaning a general element of the subset is described by specifying its components by formulas. The rule of thumb is: if each component is a linear expression of variables with constant term zero then it is a **subspace** . If some componenet is not linear then it will not be a subspace and we usually show that it is not closed under **scalar multiplication**.

When the components are linear but there are non-zero constant terms then more analysis is necessary. The main idea is to express the set V in the form $S + \boldsymbol{p}$ where S is a **subspace** and $\boldsymbol{p}$ some vector. If $\boldsymbol{p} \in S$ then it V is a **subspace** , otherwise it is not.

Example 2.5.17. Let $V = \{v(x, y, z) = \begin{pmatrix} 2x - y + 3z \\ -x + 4y - z \\ 3x + 2z \\ -2y + 7z \end{pmatrix} : x, y, z \in \mathbb{R}\}$. Determine with a proof whether V is a **subspace** of $\mathbb{R}^4$.

Since each of the components is a linear expression of the variables x, y, z without constant term we conclude that this is a subspace. We demonstrate this in two ways. The first is a powerful method and the second is "brute force" approach

(Finessed) V consists of all elements $\begin{pmatrix} 2x-y+3z \\ -x+4y-z \\ 3x+2z \\ -2y+7z \end{pmatrix}$ for arbitrary x, y, z. Now

$$\begin{pmatrix} 2x-y+3z \\ -x+4y-z \\ 3x+2z \\ -2y+7z \end{pmatrix} = x\begin{pmatrix} 2 \\ -1 \\ -1 \\ 3 \end{pmatrix} + y\begin{pmatrix} -1 \\ 4 \\ 0 \\ -2 \end{pmatrix} + z\begin{pmatrix} 3 \\ -1 \\ 2 \\ 7 \end{pmatrix}$$

and consequently V is the **span** of the sequence $\left(\begin{pmatrix} 2 \\ -1 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 4 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ 2 \\ 7 \end{pmatrix}\right)$. It follows that V is a **subspace** of $\mathbb{R}^4$.

(Brute force) We show that V is **closed under vector addition and scalar multiplication**.

Two typical elements of V are $v(x_1, y_1, z_1) = \begin{pmatrix} 2x_1-y_1+3z_1 \\ -x_1+4y_1-z_1 \\ 3x_1+2z_1 \\ -2y_1+7z_1 \end{pmatrix}$ and $v(x_2, y_2, z_2) = \begin{pmatrix} 2x_2-y_2+3z_2 \\ -x_2+4y_2-z_2 \\ 3x_2+2z_2 \\ -2y_2+7z_2 \end{pmatrix}$.

A simple calculation shows that $v(x_1, y_1, z_1)+v(x_2, y_2, z_2) = v(x_1+x_2, y_1+y_2.z_1+z_2) \in V$.

Another calculation shows that $cv(x, y, z) = v(cx, cy, cz) \in V$. So, V is a **subspace** .

Example 2.5.18. Let $V = \{\begin{pmatrix} 2x \\ x^2 \\ -3x \end{pmatrix} : x \in \mathbb{R}\}$. Determine, with a proof, whether or not V is a **subspace** of $\mathbb{R}^3$.

Since the components are not all linear we conclude that V is not a **subspace** and try and prove this. In particular, we try to show that **closure under scalar multiplication** is violated.

Taking $x = 1$ we see that $\boldsymbol{v} = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \in V$. But then $-\boldsymbol{v} = \begin{pmatrix} -2 \\ -1 \\ 3 \end{pmatrix}$ must be in V. However, $x^2 \geq 0$ and therefore the second component can never be negative.

Example 2.5.19. Let $V = \{\boldsymbol{v}(x,y) = \begin{pmatrix} 3x-y-2 \\ x+3y-4 \\ x-y \end{pmatrix} : x,y \in \mathbb{R}\}$. Determine, with a proof, whether or not V is a **subspace** of $\mathbb{R}^3$.

The general vector can be written as $\begin{pmatrix} -2 \\ -4 \\ 0 \end{pmatrix} + x \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + y \begin{pmatrix} -1 \\ 3 \\ -1 \end{pmatrix}$ and so $V = Span\left(\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ -1 \end{pmatrix}\right) + \begin{pmatrix} -2 \\ -4 \\ 0 \end{pmatrix}$.

We need to check if the vector $\begin{pmatrix} -2 \\ -4 \\ 0 \end{pmatrix} \in Span\left(\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ -1 \end{pmatrix}\right)$. We do this by making the augmented matrix

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

By application of **Gaussian elimination** we obtain the **reduced echelon form** of this matrix which is $\begin{pmatrix} 1 & 0 & | & -1 \\ 0 & 1 & | & -1 \\ 0 & 0 & | & 0 \end{pmatrix}$. Since the last column is not a **pivot column** we conclude that $\begin{pmatrix} -2 \\ -4 \\ 0 \end{pmatrix} \in Span\left(\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ -1 \end{pmatrix}\right)$ and V is a **subspace** of $\mathbb{R}^3$.

Example 2.5.20. Determine, with a proof, if $V = \{\boldsymbol{v}(x,y) = \begin{pmatrix} x+3y \\ x+4y+1 \\ x+5y+1 \end{pmatrix} : x,y \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$. a

As in **Example** (2.5.19) we can write

$$V = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}\right) + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

So we have to determine if the **linear system** with **augmented matrix** $\begin{pmatrix} 1 & 3 & | & 0 \\ 1 & 4 & | & 1 \\ 1 & 5 & | & 1 \end{pmatrix}$ is **consistent**. However, the **reduced echelon form** of this matrix is

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

The last column is a **pivot column** , the **linear system** is **inconsistent**, and we conclude that V is not a **subspace** of $\mathbb{R}^3$.

Method 2.5.2. Given a **subspace** V of $\mathbb{R}^n$ and an **n-vector** $\boldsymbol{u}$, determine if $\boldsymbol{u} \in V$.

Typically V is either given specifically in terms of formula for the components of its elements or as a **span** of a sequence of vectors. We saw in the examples illustrating **Method** (2.5.1) if the subspace is given in terms of formulas for its components how we are able to then express it as a span of a sequence of vectors. Therefore, we will assume that V is given this way, that is, $V = Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$. Now use **Method** (2.3.1). Thus, make the matrix A with columns $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ and augment by $\boldsymbol{u}$. Then $\boldsymbol{u}$ is in $Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ if and only if the **linear system** with **matrix** A is **consistent**. In turn, this is true if and only if the last column of A is not a **pivot column** which can be determined by using **Gaussian elimination** to find an **echelon form** of A.

Example 2.5.21. Determine if the **3-vector** $\boldsymbol{u} = \begin{pmatrix} 5 \\ -2 \\ -3 \end{pmatrix}$ is in $Span\left(\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \\ -8 \end{pmatrix}\right) = V$.

We form the matrix $\begin{pmatrix} 1 & 3 & | & 5 \\ 2 & 5 & | & -2 \\ -3 & -8 & | & -3 \end{pmatrix}$, use **Gaussian elimination** to obtain an **echelon form** and determine if the last column is a **pivot column** .

$\to \begin{pmatrix} 1 & 3 & | & 5 \\ 0 & -1 & | & -12 \\ 0 & 1 & | & 12 \end{pmatrix} \to \begin{pmatrix} 1 & 3 & | & 5 \\ 0 & 1 & | & 12 \\ 0 & 0 & | & 0 \end{pmatrix}$. The last column is not a **pivot column** so, yes, the vector $\boldsymbol{u} \in V$.

Method 2.5.3. Given a **subspace** V of $\mathbb{R}^n$ and a sequence $\mathcal{B} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$, determine if $\mathcal{B}$ is a basis of V.

As mentioned in **Method** (2.5.2) we may assume that V is expressed as a **span** of vectors, $V = Span(\boldsymbol{u}_1, \dots, \boldsymbol{u}_m)$. We need to establish two things: (i) that $\mathcal{B}$ is **linearly independent**, and (ii) $\mathcal{B}$ is a **spans** V.

To do (i) use **Method** (2.4.1): Make the matrix A with columns the vectors of the sequence $\mathcal{B}$, use **Gaussian elimination** to obtain an **echelon form** and determine if every column of A is a **pivot column** .

To establish (ii) we need to show that each of the vectors $\boldsymbol{u}_j$ is a **linear combination** of $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$. Each of these is done by **Method** (2.3.1): Adjoint $\boldsymbol{u}_j$ to the matrix A and determine if the last column is a **pivot column** . If not, then $\boldsymbol{u}_j \in Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$. However, we can do them all at once by augmenting the matrix with all the $\boldsymbol{u}_j$ at once. In fact, we can do both parts at once. Make the matrix $[A|\boldsymbol{u}_1 \ \dots \boldsymbol{u}_m]$ and apply **Gaussian elimination** to obtain an **echelon form** . If some column on the left hand side is not a **pivot column** then $\mathcal{B}$ is **linearly dependent**and not a basis of V. If some column on the right hand side is a pivot column then $\mathcal{B}$ is not a **spanning sequence** of V and hence not a basis. However, if every column on the left hand side is a pivot column and no column on the right hand side is a pivot column then $\mathcal{B}$ is a basis of V.

Example 2.5.22. Let V be the **subspace** of $\mathbb{R}^4$ spanned by the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ where

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -3 \end{pmatrix}.$$

Let $\mathcal{B} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ where

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ -3 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 3 \\ 3 \\ -8 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 2 \\ 5 \\ 2 \\ -7 \end{pmatrix}\}$$

Determine if $\mathcal{B}$ is a basis for V.

So, we make the matrix $\begin{pmatrix} 1 & 2 & 2 & | & 1 & 0 & 0 \\ 2 & 3 & 5 & | & 0 & 1 & 0 \\ 1 & 3 & 2 & | & 0 & 0 & 1 \\ -3 & -8 & -7 & | & 2 & -1 & -3 \end{pmatrix}$ and we use **Gaussian elimination** to get an **echelon form** . We show the reduction:

$$\to \begin{pmatrix} 1 & 2 & 2 & | & 1 & 0 & 0 \\ 0 & -1 & 1 & | & -2 & 1 & 0 \\ 0 & 1 & 0 & | & -1 & 0 & 1 \\ 0 & -2 & -1 & | & 5 & -1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 2 & | & 1 & 0 & 0 \\ 0 & 1 & 0 & | & -1 & 0 & 1 \\ 0 & -1 & 1 & | & -2 & 1 & 0 \\ 0 & -2 & -1 & | & 5 & -1 & -3 \end{pmatrix} \to$$

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

We see from this that $\mathcal{B}$ is **linearly independent** and that each of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ is in $Span(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ and therefore $\mathcal{B}$ is a basis for V.

Example 2.5.23. Let V be the **subspace** of $\mathbb{R}^4$ which is **spanned** by the sequence $\left(\begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ -2 \\ -3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \\ -4 \end{pmatrix} \right)$. Let $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 0 \\ -4 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -4 \\ 0 \\ 3 \end{pmatrix} \right)$. Determine if $\mathcal{B}$ is a basis of V.

We make the following matrix:

$$\begin{pmatrix} 1 & 1 & | & 1 & 2 & 0 & 1 \\ 0 & -4 & | & 1 & 3 & 1 & 2 \\ -4 & 0 & | & -2 & -2 & 2 & 1 \\ 3 & 3 & | & 0 & -3 & -3 & -4 \end{pmatrix}$$

After applying **Gaussian elimination** we obtain the following **echelon form** :

$$\begin{pmatrix} 1 & 1 & | & 1 & 2 & 0 & 1 \\ 0 & 1 & | & -\frac{1}{4} & -\frac{3}{4} & -\frac{1}{4} & -\frac{1}{2} \\ 0 & 0 & | & 1 & 3 & 1 & \frac{7}{3} \\ 0 & 0 & | & 0 & 0 & 0 & 0 \end{pmatrix}$$

We conclude from this that $\mathcal{B}$ is **linearly independent** but does not **span** V and therefore $\mathcal{B}$ is not a **basis** of V.

Example 2.5.24. Let V be the **subspace** of $\mathbb{R}^4$ which is **spanned** by the sequence

$\left(\begin{pmatrix}1\\1\\2\\-4\end{pmatrix}, \begin{pmatrix}2\\-1\\1\\-2\end{pmatrix}, \begin{pmatrix}1\\0\\1\\-2\end{pmatrix}, \begin{pmatrix}1\\-2\\-1\\2\end{pmatrix}\right)$. Let $\mathcal{B}$ be the sequence $\left(\begin{pmatrix}1\\-1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\\-2\end{pmatrix}, \begin{pmatrix}3\\-1\\2\\-4\end{pmatrix}\right)$.

Determine if $\mathcal{B}$ is a **basis** of V.

We make the following matrix:

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

After applying **Gaussian elimination** to this matrix we obtain the following **echelon form** :

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

We see from the **echelon form** that $\mathcal{B}$ is a **spanning sequence** of V but is **linearly dependent**and therefore not a **basis** of V.

Method 2.5.4. 5. Given a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ for a **subspace** V of $\mathbb{R}^n$ and an **n-vector** $\boldsymbol{u} \in V$ determine the **coordinate vector** , $[\boldsymbol{u}]_{\mathcal{B}}$, of $\boldsymbol{u}$ with respect to the basis $\mathcal{B}$.

The **coordinate vector** of $\boldsymbol{u}$ with respect to the **basis** $\mathcal{B}$ of V is the unique **solution** to the **linear system** with **augmented matrix** $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_k \mid \boldsymbol{u})$.

Example 2.5.25. Let $V = \{\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix} | x_1 + x_2 + x_3 - x_4 = 0\}$, a **subspace** of $\mathbb{R}^4$.

The sequence $\mathcal{B} = \left(\begin{pmatrix}1\\1\\1\\3\end{pmatrix}, \begin{pmatrix}1\\2\\3\\6\end{pmatrix}, \begin{pmatrix}2\\1\\1\\4\end{pmatrix}\right)$ is a **basis** of V. Clearly the **4-vector** $\boldsymbol{u} =$

$\begin{pmatrix} 4 \\ -3 \\ 1 \\ 2 \end{pmatrix}$ belongs to V. Find the **coordinate vector** of $\boldsymbol{u}$ with respect to $\mathcal{B}$.

Make the matrix $\begin{pmatrix} 1 & 1 & 2 & | & 4 \\ 1 & 2 & 1 & | & -3 \\ 1 & 3 & 1 & | & 1 \\ 3 & 6 & 4 & | & 2 \end{pmatrix}$.

Using **Gaussian elimination** we obtain the **reduced echelon form** of this matrix. The last column will give the **coordinate vector** :

$$\begin{pmatrix} 1 & 0 & 0 & | & -22 \\ 0 & 1 & 0 & | & 4 \\ 0 & 0 & 1 & | & 11 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

Thus, the **coordinate vector** of $\boldsymbol{u}$ is $[\boldsymbol{u}]_{\mathcal{B}} = \begin{pmatrix} -22 \\ 4 \\ 11 \end{pmatrix}$.

Exercises

In 1 - 4 determine if the given set is a **subspace** of the approppropriate $\mathbb{R}^n$ and justify your conclusion. See **Method** (2.5.1).

1. Set $\boldsymbol{v}(x, y) = \begin{pmatrix} 3x - 2y \\ -2x + y \\ x + y \end{pmatrix}$ and $V = \{\boldsymbol{v}(x, y) : x, y \in \mathbb{R}\}$.

2. Set $\boldsymbol{v}(x, y) = \begin{pmatrix} 2x + 3y \\ 3x + 4y + 1 \\ x + 3y - 3 \end{pmatrix}$ and $V = \{\boldsymbol{v}(x, y) : x, y \in \mathbb{R}\}$.

3. Set $\boldsymbol{v}(x) = \begin{pmatrix} x^2 \\ x^3 \end{pmatrix}$ and $V = \{\boldsymbol{v}(x) : x \in \mathbb{R}\}$.

4. $\boldsymbol{v}(x, y, z) = \begin{pmatrix} x + 2y + 3z \\ x - 2y - z \\ x + y - 4z \\ -3x - y - 2z + 1 \end{pmatrix}$ and $V = \{\boldsymbol{v}(x, y, z) : x, y, z \in \mathbb{R}\}$.

In 5 - 7 determine if the the vector $\boldsymbol{u}$ is in the given **subspace** V. See **Method** (2.5.2).

5. $V = \{\begin{pmatrix} x-y \\ -x+2y \\ 5x-3y \end{pmatrix} : x, y \in \mathbb{R}\}, \boldsymbol{u} = \begin{pmatrix} 2 \\ -5 \\ 2 \end{pmatrix}.$

6. $V = Span\{\begin{pmatrix} 2 \\ -3 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} -3 \\ 4 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ -4 \\ 5 \\ -1 \end{pmatrix}\}, \boldsymbol{u} = \begin{pmatrix} 3 \\ -3 \\ 5 \\ 2 \end{pmatrix}.$

7. V same as 6, $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ -3 \\ 2 \end{pmatrix}.$

In 8 - 11 determine if the sequence $\mathcal{B}$ is a **basis** for the given **subspace** V. Note that you have to first check that the sequence consists of vectors from V. See **Method** (2.5.3).

8. $V = Span\left(\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -19 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -8 \end{pmatrix}\right), \mathcal{B} = \left(\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}, \begin{pmatrix} 2 \\ 7 \\ -1 \end{pmatrix}\right).$

9. $V = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}\right), \mathcal{B} = \left(\begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\right).$

10. $V = Span\left(\begin{pmatrix} 4 \\ 3 \\ 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 5 \\ 3 \\ 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 1 \\ 5 \end{pmatrix}\right), \mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \\ 3 \end{pmatrix}\right).$

11. Let $V = \{\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} | x_1 - x_2 + x_3 - x_4 = 0\}, \mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}\right).$

In 12-14 find the **coordinate vector** of $\boldsymbol{v}$ with respect to the given **basis** $\mathcal{B}$ for the **subspace** V of $\mathbb{R}^n$.

12. $V = Span\ \mathcal{B}, \mathcal{B} = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}\right), \boldsymbol{v} = \begin{pmatrix} 2 \\ 3 \\ -1 \\ 0 \end{pmatrix}.$

13. $V = Span\ \mathcal{B}, \mathcal{B} = \left(\begin{pmatrix} 3 \\ -3 \\ 7 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 6 \\ -1 \end{pmatrix}\right), \boldsymbol{v} = \begin{pmatrix} 5 \\ -5 \\ 1 \\ 1 \end{pmatrix}.$

14. $V = Span\ \mathcal{B}, \mathcal{B} = \left(\begin{pmatrix} 1 \\ 3 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ -1 \\ -2 \end{pmatrix} \right), \boldsymbol{v} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 6 \end{pmatrix}$.

In exercises 15- 18 answer true or false and give an explanation.

15. $\mathbb{R}^2$ is a **subspace** of $\mathbb{R}^3$.

16. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a **linearly independent** sequence of vectors from $\mathbb{R}^3$ then it is a **basis** of $\mathbb{R}^3$.

17. Every **subspace** of $\mathbb{R}^4$ has **dimension** less than four.

18. If V is a **subspace** of $\mathbb{R}^4$, $V \neq \mathbb{R}^4$ and $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ a sequence of vectors from V then $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent** .

Challenge Exercises (Problems)

1. Prove that every **linearly independent** sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of **n-vectors** is a **basis** of some **subspace** of $\mathbb{R}^n$.

2. Assume $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is a **linearly dependent** sequence of **n-vectors** . Prove that the **dimension** of $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is less than k.

3. Assume U and V are **subspaces** of $\mathbb{R}^4$ each with **dimension** two. Further assume that $U \cap V = \{\boldsymbol{0}_4\}$. Prove that for every vector $\boldsymbol{x}$ in $\mathbb{R}^4$ there exists vectors $\boldsymbol{u} \in U$ and $\boldsymbol{v} \in V$ such that $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$.

Quiz Solutions

1. The **reduced echelon form** of the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3) = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 1 & 0 & 3 \\ -3 & -5 & -6 \end{pmatrix}$, computed using **Gaussian elimination** is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Since each column is a **pivot column** the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**. Not right then see **Method** (2.4.1).

2. The **reduced echelon form** of the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3\ \boldsymbol{v}_4) = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 1 & 3 & 1 & -1 \\ 1 & 0 & 3 & 2 \\ -3 & -5 & -6 & -2 \end{pmatrix}$,

computed using **Gaussian elimination** , is $R = \begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. Since the last column is not a **pivot column** the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly dependent** .

The **homogeneous linear system** which has R as its **coefficient matrix** is

$$\begin{array}{ccccccc} x_1 & & & - & x_4 & = & 0 \\ & x_2 & & + & x_4 & = & 0 \\ & & x_3 & + & 3x_4 & = & 0 \end{array}$$

There are three **leading variable** (x_1, x_2, x_3) and one **leading variable** (x_4). We set $x_4 = t$ and solve for all variables in terms of the parameter t. We get $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 7t \\ -t \\ -3t \\ t \end{pmatrix}$. Taking $t = 1$ we get the specific **non-trivial dependence relation**

$$7\boldsymbol{v}_1 - \boldsymbol{v}_2 - 3\boldsymbol{v}_3 + \boldsymbol{v}_4 = \boldsymbol{0}_4.$$

Not right, see **Method** (2.4.2).

3. $\boldsymbol{v}_4 = -7\boldsymbol{v}_1 + \boldsymbol{v}_2 - 3\boldsymbol{v}_3$.

4. The **reduced echelon form** of the matrix $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3\ \boldsymbol{v}_5) = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 1 & 3 & 1 & -1 \\ 1 & 0 & 3 & 4 \\ -3 & -5 & -6 & -3 \end{pmatrix}$, computed using **Gaussian elimination** , is $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Since every column contains a **pivot positions** it follows that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_5)$ is a **spanning sequence** of $\mathbb{R}^4$.

5. To express and $\boldsymbol{e}_i$ is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ we find the unique **solution** to the **linear system** with **augmented matrix** $[A|\boldsymbol{e}_i]$. Using **Gaussian elimination** we compute the **reduced echelon form** which will have the form $[I_4|\boldsymbol{u}_i]$ where $I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and $\boldsymbol{u}_i$ is a **4-vector** which gives the solution. However, instead of

doing this as four separate problems we can do them all at once by adjoining the four vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4$:

Using **Gaussian elimination** we compute the **reduced echelon form** of the matrix $[A|\boldsymbol{e}_1\ \boldsymbol{e}_2\ \boldsymbol{e}_3\ \boldsymbol{e}_4] = \begin{pmatrix} 1 & 2 & 2 & 1 & | & 1 & 0 & 0 & 0 \\ 1 & 3 & 1 & -1 & | & 0 & 1 & 0 & 0 \\ 1 & 0 & 3 & 4 & | & 0 & 0 & 1 & 0 \\ -3 & -5 & -6 & -3 & | & 0 & 0 & 0 & 1 \end{pmatrix}$. The matrix we obtain is $\begin{pmatrix} 1 & 0 & 0 & 0 & | & -10 & 5 & 3 & -1 \\ 0 & 1 & 0 & 0 & | & 3 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & |2 & -3 & -2 & -1 & \\ 0 & 0 & 0 & 1 & | & 1 & 1 & 1 & 1 \end{pmatrix}$. We conclude that

$$\boldsymbol{e}_1 = (-10)\boldsymbol{v}_1 + 3\boldsymbol{v}_2 + 2\boldsymbol{v}_3 + \boldsymbol{v}_5$$

$$\boldsymbol{e}_2 = 5\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + (-3)\boldsymbol{v}_3 + \boldsymbol{v}_5$$

$$\boldsymbol{e}_3 = 3\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + (-2)\boldsymbol{v}_3 + (-1)\boldsymbol{v}_5$$

$$\boldsymbol{e}_4 = \boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3 + \boldsymbol{v}_5$$

Not right, see **Method** (2.2.2).

6. Since the **reduced echelon form** of the matrix A is I_4 every column of A is a **pivot column** and consequently the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_5)$ is **linearly independent**.

Not right, see **Method** (2.4.1).

2.6. The Dot Product in $\mathbb{R}^n$

In this section we introduce the dot product which will allow us to define a length of a vector and the angle between vectors. We further prove some theorems from classical geometry, in particular, the Pythagorean theorem and its converse. We also introduce the notion of orthogonal projection.

Am I Ready for This Material

In order to follow this section and master its ideas and methods you need to be familiar with many previously defined concepts. **Those that are used extensively in this section are**:

solution of a linear system

addition of $n-$vectors

scalar multiplication of an $n-$vector

Quiz

1. Compute the **sum** of the given **n-vectors** .

a) $\begin{pmatrix} 1 \\ -3 \\ 2 \\ -5 \end{pmatrix}, \begin{pmatrix} -9 \\ 5 \\ 2 \\ -4 \end{pmatrix}.$

b) $\begin{pmatrix} 1 \\ -3 \\ 6 \end{pmatrix}, \begin{pmatrix} 4 \\ -2 \\ -2 \end{pmatrix}.$

c) $\begin{pmatrix} 2 \\ -3 \\ 4 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$

2. Compute the **scalar product** $c\boldsymbol{v}$ for the given vector $\boldsymbol{v}$ and scalar c.

a) $c = \frac{1}{2}, \boldsymbol{v} = \begin{pmatrix} -4 \\ 0 \\ 8 \\ 6 \end{pmatrix}.$

b) $c = -3, \boldsymbol{v} = \begin{pmatrix} -1 \\ 2 \\ 5 \\ -2 \end{pmatrix}.$

c) $c = -1, \boldsymbol{v} = \begin{pmatrix} -2 \\ 3 \\ -1 \\ 6 \\ -5 \end{pmatrix}.$

3. Find the **solution set** to the following **homogeneous linear system**:

$$\begin{array}{ccccccccc} 2x_1 & + & 3x_2 & - & x_3 & + & 2x_4 & = & 0 \\ 3x_1 & + & 4x_2 & - & 3x_3 & + & 4x_4 & = & 0 \\ & & x_2 & + & 3x_3 & - & 2x_4 & = & 0 \end{array}$$

4. Find the **solution set** to the following **inhomogeneous linear system**:

$$\begin{array}{ccccccccc} x_1 & + & 2x_2 & - & x_3 & - & 2x_4 & = & 0 \\ 3x_1 & + & 7x_2 & & & - & 4x_4 & = & 3 \\ 2x_1 & + & 3x_2 & - & 5x_3 & - & 4x_4 & = & -3 \end{array}.$$

Quiz Solutions

New Concepts

In this section we introduce a number of concepts that provide the setting for ordinary (affine) geometry though we will not develop this extensively here. These concepts are:

dot product

orthogonal or perpendicular vectors

norm, length, magnitude of a vector

distance between two vectors

unit vector

angle between two vectors

orthogonal complement of a subspace

projection of a vector onto the span of a given vector

projection of a vector orthogonal to a given vector

Theory (Why It Works)

In this section we introduce the **dot product** of **n-vectors**. The **dot product** allows us to define the length of a vector, the distance between two vectors, and the angle between two vectors. The dot product is the foundation of Euclidean geometry.

Definition 2.33. Let $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ be two **n-vectors**. Then the ***dot product*** of $\boldsymbol{u}$ and $\boldsymbol{v}$ is given by $\boldsymbol{u} \cdot \boldsymbol{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$.

In words, you compute the **dot product** of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ by multiplying the corresponding components of $\boldsymbol{u}$ and $\boldsymbol{v}$ and add up.

Example 2.6.1. Find then **dot product** of $(\boldsymbol{u}, \boldsymbol{v}), (\boldsymbol{u}, \boldsymbol{w})$ and $(\boldsymbol{v}, \boldsymbol{w})$ where $\boldsymbol{u} = \begin{pmatrix} 1 \\ -2 \\ 2 \\ -3 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} -2 \\ 1 \\ 3 \\ 4 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 2 \end{pmatrix}$.

$$\boldsymbol{u} \cdot \boldsymbol{v} = (1)(-2) + (-2)(1) + (2)(3) + (-3)(4) = -10,$$

$$\boldsymbol{u} \cdot \boldsymbol{w} = (1)(4) + (-2)(3) + (2)(4) + (-3)(2) = 0,$$

$$\boldsymbol{v} \cdot \boldsymbol{w} = (-2)(4) + (1)(3) + (3)(4) + (4)(2) = 15.$$

Definition 2.34. When the **dot product** of two vectors $\boldsymbol{u}, \boldsymbol{v}$ is zero we say that $\boldsymbol{u}, \boldsymbol{v}$ are ***perpendicular*** or ***orthogonal***. When $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal we write $\boldsymbol{u} \perp \boldsymbol{v}$.

We explore some of the properties of the **dot product**

Theorem 2.6.1. *Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be $n-$vectors and c any scalar. Then the following hold:*

1. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$. **Symmetry**
2. $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$ **Additivity**
3. $c(\boldsymbol{u} \cdot \boldsymbol{v}) = (c\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (c\boldsymbol{v})$.
4. $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$. $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ *if and only if* $\boldsymbol{u} = \boldsymbol{0}$. **Positive definite**

Proof. We do the proof in the case that $n = 3$. The general case is proved in exactly the same way:

1. $\boldsymbol{u} \cdot \boldsymbol{v} = u_1v_1 + u_2v_2 + u_3v_3$ and $\boldsymbol{v} \cdot \boldsymbol{u} = v_1u_1 + v_2u_2 + v_3u_3$ but $u_iv_i = v_iu_i$ for each $i = 1, 2, 3$ since this is multiplication of real numbers.

2. $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = u_1(v_1 + w_1) + u_2(v_2 + w_2) + u_3(v_3 + w_3) =$

$(u_1v_1 + u_1w_1) + (u_2v_2 + u_2w_2) + (u_3v_3 + u_3w_3) =$

$(u_1v_1 + u_2v_2 + u_3v_3) + (u_1w_1 + u_2w_2 + u_3w_3) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$.

3. $c\boldsymbol{u} = \begin{pmatrix} cu_1 \\ cu_2 \\ cu_3 \end{pmatrix}$ so that $(c\boldsymbol{u}) \cdot \boldsymbol{v} = (cu_1)v_1 + (cu_2)v_2 + (cu_3)v_3$. By associativity for multiplication of real numbers, $(cu_i)v_i = c(u_iv_i)$ for $i = 1, 2, 3$ and therefore $(cu_1)v_1 + (cu_2)v_2 + (cu_3)v_3 = c(u_1v_1) + c(u_2v_2) + c(u_3v_3)$ and by the distributive property, $c(u_1v_1) + c(u_2v_2) + c(u_3v_3) = c[u_1v_1 + u_2v_2 + u_3v_3] = c[\boldsymbol{u} \cdot \boldsymbol{v}]$. On the other hand, $\boldsymbol{u} \cdot (c\boldsymbol{v}) = (c\boldsymbol{v}) \cdot \boldsymbol{u}$ by 1. By what we have shown, $(c\boldsymbol{v}) \cdot \boldsymbol{u} = c[\boldsymbol{v} \cdot \boldsymbol{u}]$ and by making use of the first part again we get that $c[\boldsymbol{v} \cdot \boldsymbol{u}] = c[\boldsymbol{u} \cdot \boldsymbol{v}]$.

4. If $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$, then $\boldsymbol{u} \cdot \boldsymbol{u} = u_1^2 + u_2^2 + \cdots + u_n^2 \geq 0$ and is strictly greater than zero unless $u_1 = u_2 = \cdots = u_n = 0$, that is, unless $\boldsymbol{u} = \boldsymbol{0}_n$. $\square$

The next theorem is proved using **mathematical induction** and **Theorem** (2.6.1). It will be useful when we begin to discuss linear transformations in Chapter 3.

Theorem 2.6.2. *Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m, \boldsymbol{u}$ be* **n-vectors** *and $c_1, c_2, \ldots, c_m$ be scalars. Then*

$$\boldsymbol{u} \cdot (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m) = c_1(\boldsymbol{u} \cdot \boldsymbol{v}_1) + c_2(\boldsymbol{u} \cdot \boldsymbol{v}_2) + \cdots + c_m(\boldsymbol{u} \cdot \boldsymbol{v}_m).$$

Proof. We prove this by **induction** on m. The initial case, when $m = 1$ is just part 3 of Theorem (2.6.1). We need to prove the following: If the result holds for a particular m then it also holds for $m + 1$.

So, consider a **linear combination** $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_{m+1}\boldsymbol{v}_{m+1}$. We can write this as $[c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m] + c_{m+1}\boldsymbol{v}_{m+1}$, a sum of the two vectors $(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m)$ and $c_{m+1}\boldsymbol{v}_{m+1}$. By part 2 of Theorem (2.6.1) we have

$$\boldsymbol{u} \cdot ([c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m] + c_{m+1}\boldsymbol{v}_{m+1}) =$$

$$\boldsymbol{u} \cdot (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m) + \boldsymbol{u} \cdot (c_{m+1}\boldsymbol{v}_{m+1}).$$

By the inductive hypothesis ("the result holds for m") we are assuming that

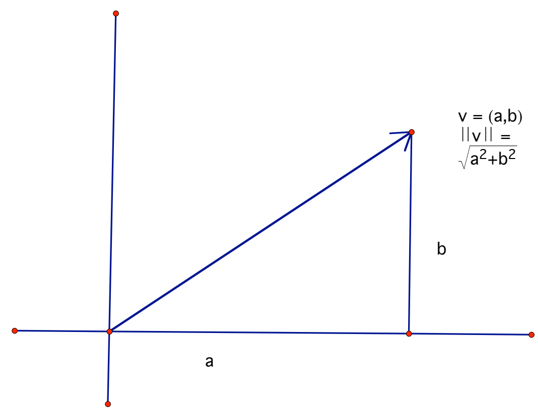


Figure 2.6.1: Length or norm of a vector in $\mathbb{R}^2$

$$\boldsymbol{u} \cdot (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m) = c_1(\boldsymbol{u} \cdot \boldsymbol{v}_1) + c_2(\boldsymbol{u} \cdot \boldsymbol{v}_2) + \cdots + c_m(\boldsymbol{u} \cdot \boldsymbol{v}_m).$$

By part 3 of Theorem (2.6.1), $\boldsymbol{u} \cdot (c_{m+1}\boldsymbol{v}_{m+1}) = c_{m+1}[\boldsymbol{u} \cdot \boldsymbol{v}_{m+1}]$. Putting this all together we get

$$\boldsymbol{u} \cdot ([c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m] + c_{m+1}\boldsymbol{v}_{m+1}) =$$

$$[c_1(\boldsymbol{u} \cdot \boldsymbol{v}_1) + c_2(\boldsymbol{u} \cdot \boldsymbol{v}_2) + \cdots + c_m(\boldsymbol{u} \cdot \boldsymbol{v}_m)] + c_{m+1}(\boldsymbol{u} \cdot \boldsymbol{v}_{m+1}) =$$

$$c_1(\boldsymbol{u} \cdot \boldsymbol{v}_1) + c_2(\boldsymbol{u} \cdot \boldsymbol{v}_2) + \cdots + c_m(\boldsymbol{u} \cdot \boldsymbol{v}_m) + c_{m+1}(\boldsymbol{u} \cdot \boldsymbol{v}_{m+1})$$

by associativity of vector addition. □

In the next definition we make use of the **dot product** to define the length of a vector:

Definition 2.35. ***Norm, length, magnitude of a vector in*** $\mathbb{R}^n$

Let $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ be an **n-vector**. Then the ***norm, length,magnitude*** of the vector $\boldsymbol{u}$, denoted by $\| \boldsymbol{u} \|$ is is

$$\sqrt{\boldsymbol{u} \cdot \boldsymbol{u}} = \sqrt{u_1^2 + u_2^2 + \ldots u_n^2}.$$

The **length (norm or magnitude)** of a vector is always defined since $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$ by part 4 of **Theorem** (2.6.1) and therefore we can always take a square root.

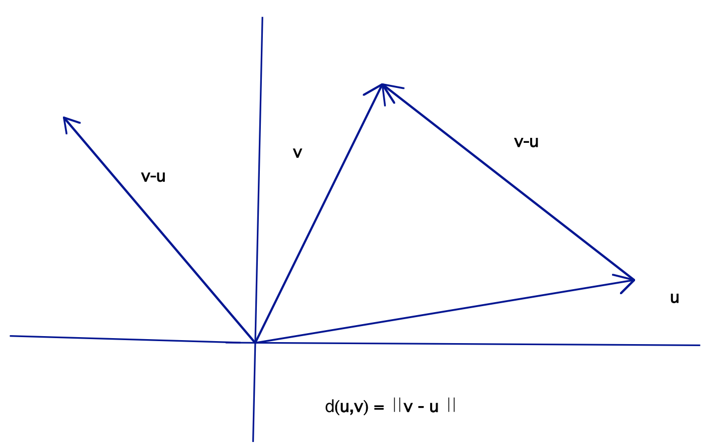
Example 2.6.2. Find the **length** of the vector $\begin{pmatrix} 1 \\ -2 \\ 2 \\ -4 \end{pmatrix}$.

$$\left\| \begin{pmatrix} 1 \\ -2 \\ 2 \\ -4 \end{pmatrix} \right\| = \sqrt{(1)^2 + (-2)^2 + (2)^2 + (-4)^2} = \sqrt{1+4+4+16} = \sqrt{25} = 5.$$

We use the **norm** of an **n-vector** to introduce the concept of distance between vectors:

Definition 2.36. For **n-vectors** $\boldsymbol{u}, \boldsymbol{v}$ the ***Euclidean distance*** or just ***distance***, between $\boldsymbol{u}$ and $\boldsymbol{v}$ is given by $dist(\boldsymbol{u}, \boldsymbol{v}) = \| \boldsymbol{u} - \boldsymbol{v} \|$.

Figure 2.6.2: Distance between two vectors in $\mathbb{R}^2$



Remark 2.10. This notion of distance is identical with the distance between points in $\mathbb{R}^2$ and $\mathbb{R}^3$ given by the usual distance formula.

Example 2.6.3. Find the **distance** between the vectors $\begin{pmatrix} 1 \\ -2 \\ 4 \\ -3 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

The **distance** is

$$\parallel \begin{pmatrix} 1 \\ -2 \\ 4 \\ -3 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \parallel =$$

$$\parallel \begin{pmatrix} 0 \\ -3 \\ 3 \\ -4 \end{pmatrix} \parallel = \sqrt{0^2 + (-3)^2 + 3^2 + (-4)^2} =$$

$$\sqrt{9+9+16} = \sqrt{34}.$$

Remark 2.11. If $\boldsymbol{u}$ is a vector and c is a scalar then $\parallel c\boldsymbol{v} \parallel = |c| \parallel \boldsymbol{u} \parallel$.

A consequence of this is the following:

Theorem 2.6.3. *Let $\mathbf{u}$ be a non-zero vector. Then the* **norm** *of $\frac{1}{\|\mathbf{u}\|}\mathbf{u}$ is 1.*

Proof. $\parallel \frac{1}{\|\boldsymbol{u}\|}\boldsymbol{u} \parallel = \frac{1}{\|\boldsymbol{u}\|} | \parallel \boldsymbol{u} \parallel = \frac{1}{\|\boldsymbol{u}\|} \parallel \boldsymbol{u} \parallel = 1.$ □

Vectors of **norm (length)** equal to one are very important; for further reference we give them a name:

Definition 2.37. A vector of **norm (length)** one is called a ***unit vector***. When we divide a non-zero vector by its **norm** we say we are ***normalizing*** and the vector so obtained is said to be ***a unit vector in the direction of*** $\boldsymbol{u}$.

Our next goal is to prove several theorems which will allow us to introduce angles between vectors in $\mathbb{R}^n$ in a rigorous way.

Theorem 2.6.4. *Let $\boldsymbol{u}, \boldsymbol{v}$ be vectors in $\mathbb{R}^n$. Then*

$$|\boldsymbol{u} \centerdot \boldsymbol{v}| \leq \| \boldsymbol{u} \| \cdot \| \boldsymbol{v} \| \tag{2.17}$$

with equality if and only if the sequence $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly dependent**. *This result is referred to as the* **Cauchy-Schwartz inequality**.

Proof. If $\boldsymbol{u}$ or $\boldsymbol{v}$ is the zero vector then both sides in (2.17) are zero and therefore we may assume that $\boldsymbol{u}, \boldsymbol{v} \neq \mathbf{0}_n$.

Let λ represent a real number and consider the **linear combination** $\lambda\boldsymbol{u} + \boldsymbol{v}$. Making use of the properties of the dot product, **Theorem** (2.6.1), we have

$$(\lambda\boldsymbol{u} + \boldsymbol{v}) \centerdot (\lambda\boldsymbol{u} + \boldsymbol{v}) = (\lambda\boldsymbol{u} + \boldsymbol{v}) \centerdot (\lambda\boldsymbol{u}) + (\lambda\boldsymbol{u} + \boldsymbol{v}) \centerdot \boldsymbol{v} \text{ by additivity}$$

$$= (\lambda\boldsymbol{u}) \centerdot (\lambda\boldsymbol{u}) + \boldsymbol{v} \centerdot (\lambda\boldsymbol{u}) + (\lambda\boldsymbol{u}) \centerdot \boldsymbol{v} + \boldsymbol{v} \centerdot \boldsymbol{v} \text{ again by additivity}$$

$$= \lambda^2\boldsymbol{u} \centerdot \boldsymbol{u} + \lambda\boldsymbol{u} \centerdot \boldsymbol{v} + \lambda\boldsymbol{v} \centerdot \boldsymbol{u} + \boldsymbol{v} \centerdot \boldsymbol{v} \text{ by the scalar property}$$

$$= \lambda^2\boldsymbol{u} \centerdot \boldsymbol{u} + 2\lambda\boldsymbol{u} \centerdot \boldsymbol{v} + \boldsymbol{v} \centerdot \boldsymbol{v} \text{ by symmetry.}$$

Let us set $\boldsymbol{u} \centerdot \boldsymbol{u} = A, 2\boldsymbol{u} \centerdot \boldsymbol{v} = B$ and $\boldsymbol{v} \centerdot \boldsymbol{v} = C$. With these assignments we obtain

$$(\lambda\boldsymbol{u} + \boldsymbol{v}) \centerdot (\lambda\boldsymbol{u} + \boldsymbol{v}) = A\lambda^2 + B\lambda + C \tag{2.18}$$

Note also by the **positive definiteness** of the **dot product** that $(\lambda\boldsymbol{u} + \boldsymbol{v}) \centerdot (\lambda\boldsymbol{u} + \boldsymbol{v}) = A\lambda^2 + B\lambda + C \geq 0$ with equality if and only if $\lambda\boldsymbol{u} + \boldsymbol{v} = \mathbf{0}_n$. However, $\lambda\boldsymbol{u} + \boldsymbol{v} = \mathbf{0}_n$ if and only if $\boldsymbol{v}$ is a multiple of $\boldsymbol{u}$, if and only if $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly dependent**. In addition, when the two vectors are **linearly independent** the roots must be non-real since the quadratic expression $A\lambda^2 + B\lambda + C$ is positive for every λ.

We may consider $A\lambda^2 + B\lambda + C = 0$ to be a quadratic equation in λ. Before preceeding to the conclusion of the proof we recall a some facts about quadratic equations.

Suppose

$$ax^2 + bx + c = 0 \tag{2.19}$$

is a quadratic equation with real coefficients and $a \neq 0$. The roots of equation (2.19) are given by the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{2.20}$$

We see from (2.20) that $\Delta = b^2 - 4ac > 0$ (Δ is called the ***discriminant*** of the equation) if and only if there are two distinct real roots; $\Delta = 0$ if and only if there is a single real root; and $\Delta < 0$ if and only if there are no real roots.

Returning to $A\lambda^2 + B\lambda + C = 0$ we have two possibilities: i) $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly dependent** and there is a unique solution; or ii) $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly independent** and there is no real solution.

i) In this case, since there is a single solution, the discriminant of the quadratic equation must be zero and therefore $B^2 - 4AC = 0$. Then $B^2 = 4AC$. Recalling that $A = \boldsymbol{u} \cdot \boldsymbol{u} = \| \boldsymbol{u} \|^2, B = 2\boldsymbol{u} \cdot \boldsymbol{v}, C = \boldsymbol{v} \cdot \boldsymbol{v} = \| \boldsymbol{v} \|^2$ we have that

$$(\boldsymbol{u} \cdot \boldsymbol{v})^2 = \| \boldsymbol{u} \|^2 \| \boldsymbol{v} \|^2$$

Taking square roots gives

$$|\boldsymbol{u} \cdot \boldsymbol{v}| = \| \boldsymbol{u} \| \| \boldsymbol{v} \|$$

ii) In this case the quadratic equation $A\lambda^2 + 2B\lambda + C = 0$ has no real roots and consequently the discriminant $\Delta = B^2 - 4AC < 0$. It follows from this that

$$B^2 < 4AC$$

and we conclude that

$$(\boldsymbol{u} \cdot \boldsymbol{v})^2 < \| \boldsymbol{u} \|^2 \| \boldsymbol{v} \|^2$$

Again, take square roots to get

$$|\boldsymbol{u} \cdot \boldsymbol{v}| < \| \boldsymbol{u} \| \| \boldsymbol{v} \|$$

$\square$

We can use the **Cauchy-Schwartz inquality** to prove a familiar theorem from Euclidean geometry referred to as the **triangle inequality**: The sum of the lengths of two sides of a triangle exceeds the length of the third side.

Suppose the the vertices of the triangle are $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. Then the sides are the vectors $\boldsymbol{u} = \boldsymbol{b} - \boldsymbol{a}, \boldsymbol{v} = \boldsymbol{c} - \boldsymbol{b}$ and $\boldsymbol{c} - \boldsymbol{a} = (\boldsymbol{c} - \boldsymbol{b}) + (\boldsymbol{b} - \boldsymbol{a}) = \boldsymbol{v} + \boldsymbol{u}$. Thus, to demonstrate the triangle inequality we need to show that $\| \boldsymbol{v} + \boldsymbol{u} \| < \| \boldsymbol{u} \| + \| \boldsymbol{v} \|$ when $\boldsymbol{u}, \boldsymbol{v}$ are two sides of a triangle. There is also a degenerate case, when $\boldsymbol{u}, \boldsymbol{v}$ are parallel and in the same direction and the theorem takes account of this as well.

Theorem 2.6.5. *(**Triangle Inequality**) Let $\boldsymbol{u}, \boldsymbol{v}$ be **n-vectors**. Then*

$$\| \boldsymbol{u}+\boldsymbol{v} \| \leq \| \boldsymbol{u} \| + \| \boldsymbol{v} \| . \tag{2.21}$$

When $\boldsymbol{u}, \boldsymbol{v} \neq \boldsymbol{0}_n$ we have equality if and only if there is a positive λ such that $\boldsymbol{v} = \lambda \boldsymbol{u}$.

Proof. Note that when either $\boldsymbol{u}$ or $\boldsymbol{v}$ is the zero vector there is nothing to prove and we have equality, so assume that $\boldsymbol{u}, \boldsymbol{v} \neq \boldsymbol{0}_n$.

Suppose $\boldsymbol{u}, \boldsymbol{v}$ are parallel (that is, $Span(\boldsymbol{u}) = Span(\boldsymbol{v})$ and have the same direction ($\boldsymbol{v} = \lambda\boldsymbol{u}$ with $\lambda > 0$). Then $\boldsymbol{u}+\boldsymbol{v} = (1+\lambda)\boldsymbol{u}$ has length $(1+\lambda) \| \boldsymbol{u} \|$ by **Remark** (2.11).

On the other hand,

$$\| \boldsymbol{u} \| + \| \boldsymbol{v} \| = \| \boldsymbol{u} \| + \| \lambda\boldsymbol{u} \| = \| \boldsymbol{u} \| + |\lambda| \| \boldsymbol{u} \| = \| \boldsymbol{u} \| + \lambda \| \boldsymbol{u} \| \tag{2.22}$$

the last equality since $\lambda > 0$

$$= (1+\lambda) \| \boldsymbol{u} \| \tag{2.23}$$

If $\boldsymbol{u}, \boldsymbol{v}$ are parallel but have opposite directions ($\boldsymbol{v} = \lambda\boldsymbol{u}, \lambda < 0$) then the same calculations show that $\| \boldsymbol{u}+\boldsymbol{v} \| < \| \boldsymbol{u} \| + \| \boldsymbol{v} \|$. Therefore we may assume that $(\boldsymbol{u}, \boldsymbol{v})$ is **linearly independent**, or equivalently, not parallel. Applying properties of the **dot product** we get

$\| \boldsymbol{u}+\boldsymbol{v} \|^2 = (\boldsymbol{u}+\boldsymbol{v}) \cdot (\boldsymbol{u}+\boldsymbol{v})$ by the **definition of length**;

$= (\boldsymbol{u}+\boldsymbol{v}) \cdot \boldsymbol{u} + (\boldsymbol{u}+\boldsymbol{v}) \cdot \boldsymbol{v}$ by the **additivity property** of the **dot product**;

$= \boldsymbol{u} \cdot \boldsymbol{u} + \boldsymbol{v} \cdot \boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{v}$ by the **additivity property** of the **dot product**;

$= \boldsymbol{u} \cdot \boldsymbol{u} + 2(\boldsymbol{u} \cdot \boldsymbol{v}) + \boldsymbol{v} \cdot \boldsymbol{v}$ by the **symmetry property** of the **dot product**;

$= \| \boldsymbol{u}^2 \| + 2(\boldsymbol{u} \cdot \boldsymbol{v}) + \| \boldsymbol{v} \|^2$ by the **definition of the length of a vector**.

Putting this together we have:

$$\| \boldsymbol{u}+\boldsymbol{v} \|^2 = \| \boldsymbol{u} \|^2 + 2(\boldsymbol{u} \cdot \boldsymbol{v}) + \| \boldsymbol{v} \|^2 \tag{2.24}$$

By the **Cauchy-Schwartz inequality** $\boldsymbol{u} \cdot \boldsymbol{v} \leq |\boldsymbol{u} \cdot \boldsymbol{v}| \leq \| \boldsymbol{u} \| \cdot \| \boldsymbol{v} \|$. Substituting this into (2.24) we obtain

$$\parallel \boldsymbol{u} \parallel^2 + 2(\boldsymbol{u} \cdot \boldsymbol{v}) + \parallel \boldsymbol{v} \parallel^2 \le \parallel \boldsymbol{u} \parallel^2 + 2 \parallel \boldsymbol{u} \parallel \parallel \boldsymbol{v} \parallel + \parallel \boldsymbol{v} \parallel^2 =$$

$$(\parallel \boldsymbol{u} \parallel + \parallel \boldsymbol{v} \parallel)^2$$

Thus, $\parallel \boldsymbol{u} + \boldsymbol{v} \parallel^2 \le (\parallel \boldsymbol{u} \parallel + \parallel \boldsymbol{v} \parallel)^2$. Taking square roots yields the theorem. □

As another application of the **Cauchy-Schwartz inequality** we can introduce the notion of angles between vectors:

Suppose that $\boldsymbol{u}, \boldsymbol{v}$ is a pair of (not necessarily distinct) non-zero vectors. Then

$$\frac{|\boldsymbol{u} \cdot \boldsymbol{v}|}{\parallel \boldsymbol{u} \parallel \parallel \boldsymbol{v} \parallel} \le 1 \tag{2.25}$$

from which we conclude that

$$-1 \le \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\parallel \boldsymbol{u} \parallel \parallel \boldsymbol{v} \parallel} \le 1 \tag{2.26}$$

We know from the study of the function $cos\ x$ in calculus that for any number α with $-1 \le \alpha \le 1$ there is a unique number $\theta, 0 \le \theta \le \pi$, such that $cos\ \theta = \alpha$. Denoting this by $arccos\ \alpha$ we can define an angle between vectors in $\mathbb{R}^n$ for any n:

Definition 2.38. Let $\boldsymbol{u}, \boldsymbol{v}$ be two (not necessarily distinct) vectors, not both zero. If one of the vectors is the zero vector then we define the angle between the vectors to be $\frac{\pi}{2}$. If neither $\boldsymbol{u}, \boldsymbol{v}$ is the zero vector we define the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ to be

$$arccos\ (\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\parallel \boldsymbol{u} \parallel \parallel \boldsymbol{v} \parallel})$$

Remark 2.12. 1) According to **Definition** (2.38) if $\boldsymbol{u}, \boldsymbol{v}$ are not both the zero vector then the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is $\frac{\pi}{2}$ if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthogonal** ($\boldsymbol{u} \cdot \boldsymbol{v} = 0$).

2) If θ is the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ then $\boldsymbol{u} \cdot \boldsymbol{v} = \parallel \boldsymbol{u} \parallel \parallel \boldsymbol{v} \parallel cos\ \theta$.

We can now prove a famous theorem (and its converse) familiar to many middle school students as well as nearly all high school students - the **Pythagorean theorem**:

Theorem 2.6.6. (**Pythagorean Theorem**) *Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be non-zero vectors. Then $\boldsymbol{u}$ and $\boldsymbol{v}$ are* **orthogonal** *if and only if*

$$\parallel \boldsymbol{u} + \boldsymbol{v} \parallel^2 = \parallel \boldsymbol{u} \parallel^2 + \parallel \boldsymbol{v} \parallel^2 \tag{2.27}$$

Proof. By the **definition of norm** and **Theorem** (2.6.1) we have

$$\| \boldsymbol{u}+\boldsymbol{v} \|^2 = (\boldsymbol{u}+\boldsymbol{v}) \cdot (\boldsymbol{u}+\boldsymbol{v}) =$$

$$\boldsymbol{u} \cdot \boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{u} + \boldsymbol{v} \cdot \boldsymbol{v} =$$

$$\| \boldsymbol{u} \|^2 + 2\boldsymbol{u} \cdot \boldsymbol{v} + \| \boldsymbol{v} \|^2 .$$

Then $\| \boldsymbol{u}+\boldsymbol{v} \|^2 = \| \boldsymbol{u} \|^2 + \| \boldsymbol{v} \|^2$ if and only if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthogonal**. □

When W is a **subspace** of $\mathbb{R}^n$ we will be interested in the collection of vectors which are **orthogonal** to all vectors in W. As we will see this is a **subspace** and related to **homogenous linear systems**.

Definition 2.39. Let W be a **subspace** of $\mathbb{R}^n$. The ***orthogonal complement*** to W is the collection of all vectors $\boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{x}$ is **orthogonal** to every vector in W, that is, $\boldsymbol{x} \cdot \boldsymbol{w} = 0$ for every vector $\boldsymbol{w} \in W$. We denote the orthogonal complement to W by $W^\perp$ (which is read as "W perp").

The next theorem demonstrates that $W^\perp$ is a **subspace** of $\mathbb{R}^n$. It also gives a characterization of $W^\perp$ when W is the **span** of a sequence of vectors.

Theorem 2.6.7. *1) Let W be a* **subspace** *of* $\mathbb{R}^n$*. Then $W^\perp$ is a* **subspace** *of* $\mathbb{R}^n$*.*

2) Assume that $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$. Then $\boldsymbol{x} \in W^\perp$ if and only if $\boldsymbol{w}_i \cdot \boldsymbol{x} = 0$ for each $i, 1 \le i \le k$.

Proof. 1) Clearly $\mathbf{0}_n$ is in $W^\perp$ and therefore $W^\perp$ is not empty. We need to show the following: i) If $\boldsymbol{x}_1, \boldsymbol{x}_2 \in W^\perp$ then $\boldsymbol{x}_1 + \boldsymbol{x}_2 \in W^\perp$. ii) If $\boldsymbol{x} \in W^\perp$ and $c \in \mathbb{R}$ then $c\boldsymbol{x} \in W^\perp$.

i) We need to show that for every $\boldsymbol{w} \in W, \boldsymbol{w} \cdot (\boldsymbol{x}_1 + \boldsymbol{x}_2) = 0$. However, by **additivity** of the **dot product** , $\boldsymbol{w} \cdot (\boldsymbol{x}_1 + \boldsymbol{x}_2) = \boldsymbol{w} \cdot \boldsymbol{x}_1 + \boldsymbol{w} \cdot \boldsymbol{x}_2$. Since $\boldsymbol{x}_1 \in W^\perp, \boldsymbol{w} \cdot \boldsymbol{x}_1 = 0$. Likewise, since $\boldsymbol{x}_2 \in W^\perp, \boldsymbol{w} \cdot \boldsymbol{x}_2 = 0$. Therefore, $\boldsymbol{w} \cdot (\boldsymbol{x}_1 + \boldsymbol{x}_2) = 0 + 0 = 0$ as required.

ii) We need to show that for every $\boldsymbol{w} \in W$ that $\boldsymbol{w} \cdot (c\boldsymbol{x}) = 0$. By part 3 of **Theorem** (2.6.1) $\boldsymbol{w} \cdot (c\boldsymbol{x}) = c(\boldsymbol{w} \cdot \boldsymbol{x})$. Since $\boldsymbol{x} \in W^\perp, \boldsymbol{w} \cdot \boldsymbol{x} = 0$. Thus, $\boldsymbol{w} \cdot (c\boldsymbol{x}) = 0$ as we needed to show.

2) Let W' consist of all vectors $\boldsymbol{x}$ such that $\boldsymbol{w}_i \cdot \boldsymbol{x} = 0$ for all $i, 1 \le i \le k$. We need to show that $W^\perp = W'$. To do this we need to prove that if $\boldsymbol{x} \in W^\perp$ then $\boldsymbol{x} \in W'$ and if $\boldsymbol{x} \in W'$ then $\boldsymbol{x} \in W^\perp$.

Suppose first that $\boldsymbol{x} \in W^\perp$. Then $\boldsymbol{w} \cdot \boldsymbol{x} = 0$ for every $\boldsymbol{w} \in W$. In particular, $\boldsymbol{w}_i \cdot \boldsymbol{x} = 0$ for each $i, 1 \le i \le k$ since $\boldsymbol{w}_i \in W$. Thus, $\boldsymbol{x} \in W'$.

Now assume that $\boldsymbol{x} \in W'$. We need to show for an arbitrary element $\boldsymbol{w} \in W$ that $\boldsymbol{w} \cdot \boldsymbol{x} = 0$. Since $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ if $\boldsymbol{w} \in W$ then there are scalars $c_1, c_2, \dots, c_k$ such that $\boldsymbol{w} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \cdots + c_k\boldsymbol{w}_k$.

Then $\boldsymbol{w} \cdot \boldsymbol{x} = (c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \cdots + c_k\boldsymbol{w}_k) \cdot \boldsymbol{x} = c_1(\boldsymbol{w}_1 \cdot \boldsymbol{x}) + c_2(\boldsymbol{w}_2 \cdot \boldsymbol{x}) + \cdots + c_k(\boldsymbol{w}_k \cdot \boldsymbol{x})$ by **Theorem** (2.6.2).

Since for each i, $\boldsymbol{x} \cdot \boldsymbol{w}_i = 0$, it follows that $\boldsymbol{w} \cdot \boldsymbol{x} = 0$ and $\boldsymbol{x} \in W^\perp$ as we needed to show. □

Example 2.6.4. Consider a typical **homogeneous linear system**, for example

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & a_{14}x_4 & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & a_{24}x_4 & = & 0 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & a_{34}x_4 & = & 0 \end{array} \tag{2.28}$$

The matrix of the system (2.28) is $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}$. Let $\boldsymbol{w}_i$ be the 4-vector which has the same entries as the i^{th} row of A, $\boldsymbol{w}_i = \begin{pmatrix} a_{i1} \\ a_{i2} \\ a_{i3} \\ a_{i4} \end{pmatrix}$ and set $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$. A vector $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ is a **solution** to the system (2.28) if and only if $\boldsymbol{x}$ is in the **orthogonal complement** to W.

We close this section with an introduction to the concept of an orthogonal projection. Given a nonzero vector $\boldsymbol{v}$ and an arbitrary vector $\boldsymbol{w}$ in $\mathbb{R}^n$ the objective is to express $\boldsymbol{w}$ as the sum of a multiple of $\boldsymbol{v}$ (a vector in $Span(\boldsymbol{v})$) and a vector **orthogonal** to $\boldsymbol{v}$ (in $\boldsymbol{v}^\perp$). Rather than write an unmotivated formula we derive the solution.

Thus, we want to find a scalar α such that $\boldsymbol{w}+\alpha\boldsymbol{v}$ is **orthogonal** to $\boldsymbol{v}$. Set the **dot product** of $\boldsymbol{w}+\alpha\boldsymbol{v}$ and $\boldsymbol{v}$ equal to zero and solve for α.

$$0 = (\boldsymbol{w}+\alpha\boldsymbol{v}) \cdot \boldsymbol{v} = \boldsymbol{w} \cdot \boldsymbol{v} + \alpha(\boldsymbol{v} \cdot \boldsymbol{v}) \tag{2.29}$$

Solving for α we obtain $\alpha = -\frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}$.

We check that this value of α does, indeed, make the vector $\boldsymbol{w}+\alpha\boldsymbol{v}$ **orthogonal** to $\boldsymbol{v}$:

$$[\boldsymbol{w} - \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}] \cdot \boldsymbol{v} = \boldsymbol{w}\cdot\boldsymbol{v} - \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}(\boldsymbol{v}\cdot\boldsymbol{v})$$

by the **additivity and scalar properties** of the **dot product**. Canceling the term $\boldsymbol{v}\cdot\boldsymbol{v}$ occurring in the numerator and denominator of the second term we get

$$\boldsymbol{w}\cdot\boldsymbol{v} - \boldsymbol{w}\cdot\boldsymbol{v} = 0$$

as required. We have therefore proved:

Theorem 2.6.8. *Let $\boldsymbol{v}, \boldsymbol{w}$ be* **n-vectors** *with $\boldsymbol{v} \neq \mathbf{0}_n$. Then the vector $\boldsymbol{w} - \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$ is* **orthogonal** *to $\boldsymbol{v}$ (equivalently, belongs to $\boldsymbol{v}^\perp$.)*

We give a name to the vectors $\boldsymbol{w} - \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$ and $\frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$:

Definition 2.40. Let $\boldsymbol{v}, \boldsymbol{w}$ be vectors in $\mathbb{R}^n$, $\boldsymbol{v} \neq \mathbf{0}_n$. The vector $\boldsymbol{w} - \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$ is called ***the projection of w orthogonal to v***. This is denoted by $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$. The vector $\frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$ is ***the projection of w onto v***. This is denoted by $Proj_{\boldsymbol{v}}(\boldsymbol{w})$.

The **projection of a vector orthogonal** to a nonzero vector has an interesting property which can be used to solve geometric problems. Specifically, if $\boldsymbol{v}, \boldsymbol{w}$ are vectors in $\mathbb{R}^n$ and $\boldsymbol{v}$ is nonzero then amongst all vectors of the form $\boldsymbol{w} - \alpha\boldsymbol{v}$ the vector $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$ has the shortest **length**. We state this as a theorem.

Theorem 2.6.9. *Let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ with $\boldsymbol{v}$ nonzero. Then amongst all vectors of the form $\boldsymbol{w} - \alpha\boldsymbol{v}$ the* **projection** *of $\boldsymbol{w}$ orthogonal to $\boldsymbol{v}$, $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w}) = \boldsymbol{w} - \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$, has the shortest* **length**.

Proof. Let α be a scalar and consider

$$(\boldsymbol{w} - \alpha\boldsymbol{v}) \cdot (\boldsymbol{w} - \alpha\boldsymbol{v}) = \ \parallel \boldsymbol{w} \parallel^2 -2\alpha(\boldsymbol{w} \cdot \boldsymbol{v}) + \alpha^2 \parallel \boldsymbol{v} \parallel^2 .$$

This is a quadratic function in α with a positive leading coefficient ($\parallel \boldsymbol{v} \parallel^2$). It then follows that the function has a minimum at $\frac{-(-2\boldsymbol{w}\cdot\boldsymbol{v})}{2\|\boldsymbol{v}\|^2} = \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}$. When $\alpha = \frac{\boldsymbol{w}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}$ it follows that $\boldsymbol{w} - \alpha\boldsymbol{v} = Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$. □

What You Can Now Do

1. For a sequence S of **n-vectors** , $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$, determine the **orthogonal complement** to $Span\ S$.

2. For a sequence of **n-vectors** , $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$, determine the collection of all vectors $\boldsymbol{x}$ which satisfy $\boldsymbol{x} \cdot \boldsymbol{v}_1 = b_1, \boldsymbol{x} \cdot \boldsymbol{v}_2 = b_2, \ldots, \boldsymbol{x} \cdot \boldsymbol{v}_k = b_k$.

3. For **n-vectors** $\boldsymbol{v}, \boldsymbol{w}$ with $\boldsymbol{v} \neq \boldsymbol{0}_n$, compute the **projection** of $\boldsymbol{w}$ on $\boldsymbol{v}$ and the **projection** of $\boldsymbol{w}$ orthogonal to $\boldsymbol{v}$.

4. In $\mathbb{R}^2$ find the shortest distance of a point p in $\mathbb{R}^2$ to a line L.

Method (How To Do It)

Method 2.6.1. For a sequence S of **n-vectors** , $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$, determine the **orthogonal complement** to $Span\ S$.

By **Theorem** (2.6.7) a vector $\boldsymbol{x}$ is in the **orthogonal complement** of $Span(S)$ if and only if $\boldsymbol{x} \cdot \boldsymbol{v}_i = 0$ for $1 \leq i \leq k$.

Write the unknown vector as $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. Each **dot product** of $\boldsymbol{x}$ and $\boldsymbol{v}_i$ gives a **homogeneous linear equation** and therefore this gives rise to a **homogeneous linear system**. The desired collection is the **solution space** to this **linear system**. Without having to do this every time you can just take the vectors as the **rows** of a matrix which is the coefficient matrix of the **homogeneous linear system**.

Example 2.6.5. Find all the vectors in $\mathbb{R}^4$ which are **orthogonal** to each of the vectors

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 3 \\ -1 \\ 4 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 5 \\ 1 \\ 5 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 4 \\ -4 \\ 7 \end{pmatrix}.$$

We form the matrix $A = \begin{pmatrix} 1 & 3 & -1 & 4 \\ 2 & 5 & 1 & 5 \\ 1 & 4 & -4 & 7 \end{pmatrix}$ and proceed to use **Gaussian elimination** to obtain the **reduced echelon form**, $R = \begin{pmatrix} 1 & 0 & 8 & -5 \\ 0 & 1 & -3 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

The **homogeneous linear system** with **coefficient matrix** equal to R is

$$\begin{array}{ccccccccc} x_1 & & & + & 8x_3 & - & 5x_4 & = & 0 \\ & & x_2 & - & 3x_3 & + & 3x_4 & = & 0 \end{array}$$

There are two **leading variables** (x_1, x_2) and two **free variables** $(x_3, x_4.)$ We assign parameters $x_3 = s, x_4 = t$ and express all the variables in terms of s and t.

Thus, $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -8s + 5t \\ 3s - 3t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} -8 \\ 3 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 5 \\ -3 \\ 0 \\ 1 \end{pmatrix}$. The **orthogonal complement** to $Span\left(\begin{pmatrix} 1 \\ 3 \\ -1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ -4 \\ 7 \end{pmatrix}\right)$ is $Span\left(\begin{pmatrix} -8 \\ 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ -3 \\ 0 \\ 1 \end{pmatrix}\right)$.

Method 2.6.2. For a sequence of **n-vectors** , $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$, determine the collection of all vectors $\boldsymbol{x}$ which satisfy

$$\boldsymbol{x} \cdot \boldsymbol{v}_1 = b_1, \boldsymbol{x} \cdot \boldsymbol{v}_2 = b_2, \dots, \boldsymbol{x} \cdot \boldsymbol{v}_k = b_k.$$

This reduces to solving an **(inhomogeneous) linear system**. If the vector $\boldsymbol{v}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$ then it is the **linear system** with **augmented matrix**

$$\begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} & | & b_1 \\ a_{12} & a_{22} & \dots & a_{n2} & | & b_2 \\ \vdots & \vdots & \vdots & \vdots & | & \vdots \\ a_{1k} & a_{2k} & \dots & a_{nk} & | & b_k \end{pmatrix}.$$

Note that the vector $\boldsymbol{v}_j$ is the j^{th} row of the coefficient matrix of the **linear system**. Now use **Method** (1.2.4) to find the **solution set** to the **linear system**.

Example 2.6.6. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 2 \\ 3 \\ -2 \end{pmatrix}$.

Find all vectors $\boldsymbol{x}$ such that $\boldsymbol{x} \cdot \boldsymbol{v}_1 = 0, \boldsymbol{x} \cdot \boldsymbol{v}_2 = 1, \boldsymbol{x} \cdot \boldsymbol{v}_3 = -1$.

This is the **solution set** to the **linear system** with **augmented matrix**

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

We use **Gaussian elimination** to obtain the **reduced echelon form** of this matrix:

$$R = \begin{pmatrix} 1 & 0 & 0 & -11 & | & -1 \\ 0 & 1 & 0 & 5 & | & 1 \\ 0 & 0 & 1 & 7 & | & 0 \end{pmatrix}.$$

The **linear system** with **augmented matrix** R is

$$\begin{array}{ccccccc} x_1 & & & - & 11x_4 & = & -1 \\ & x_2 & & + & 5x_4 & = & 1 \\ & & x_3 & + & 7x_4 & = & 0 \end{array}$$

There are three **leading variables** (x_1, x_2, x_3) and one **free variable** (x_4). We set $x_4 = t$ and solve for all the variables in terms of t:
This gives the general solution

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1+11t \\ 1-5t \\ -7t \\ t \end{pmatrix} =$$

$$\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 11 \\ -5 \\ -7 \\ 1 \end{pmatrix}.$$

Method 2.6.3. For **n-vectors** $\boldsymbol{v}, \boldsymbol{w}$ with $\boldsymbol{v} \neq \boldsymbol{0}_n$, compute the **projection** of $\boldsymbol{w}$ on $\boldsymbol{v}$ and the **projection** of $\boldsymbol{w}$ orthogonal to $\boldsymbol{v}$.

First compute $\frac{\boldsymbol{w} \cdot \boldsymbol{v}}{\boldsymbol{v} \cdot \boldsymbol{v}}$ and then multiply $\boldsymbol{v}$ by this scalar. This is the vector $Proj_{\boldsymbol{v}}(\boldsymbol{w})$.

The difference of $\boldsymbol{w}$ and $Proj_{\boldsymbol{v}}(\boldsymbol{w})$

$$\boldsymbol{w} - \frac{\boldsymbol{w} \cdot \boldsymbol{v}}{\boldsymbol{v} \cdot \boldsymbol{v}} \boldsymbol{v}$$

is $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$.

Example 2.6.7. For the vectors $\boldsymbol{v} = \begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} 5 \\ 1 \\ -6 \\ 1 \end{pmatrix}$ compute $Proj_{\boldsymbol{v}}(\boldsymbol{w})$ and $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$.

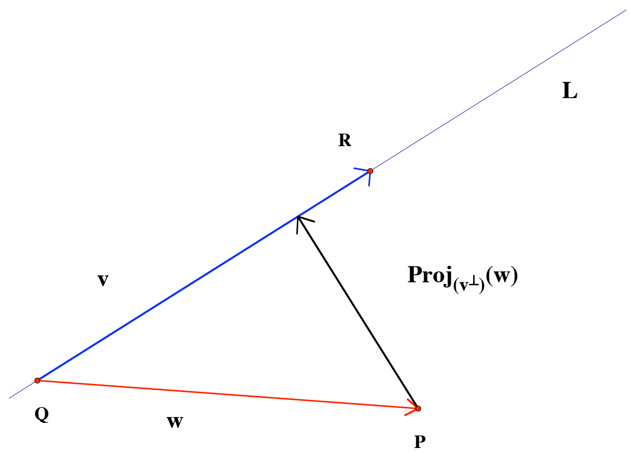
$$Proj_{\boldsymbol{v}}(\boldsymbol{w}) = \frac{\begin{pmatrix} 5 \\ 1 \\ -6 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix}}{\begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix}} \begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix} = \tfrac{27}{18} \begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix} = \tfrac{3}{2} \begin{pmatrix} 2 \\ 1 \\ -3 \\ -2 \end{pmatrix} = \begin{pmatrix} 3 \\ \frac{3}{2} \\ -\frac{9}{2} \\ -3 \end{pmatrix}.$$

$$Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w}) = \boldsymbol{w} - Proj_{\boldsymbol{v}}(\boldsymbol{w}) = \begin{pmatrix} 5 \\ 1 \\ -6 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ \frac{3}{2} \\ -\frac{9}{2} \\ -3 \end{pmatrix} = \begin{pmatrix} 2 \\ -\frac{1}{2} \\ -\frac{3}{2} \\ 4 \end{pmatrix}.$$

Method 2.6.4. In $\mathbb{R}^2$ find the shortest distance of a point P in $\mathbb{R}^2$ to a line L.

Choose any two distinct points Q and R on the line L. Set $\boldsymbol{v} = \overline{QR}$ and $\boldsymbol{w} = \overline{PQ}$. Then the desired distance is the **length** of the vector $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$.

Figure 2.6.3: Projection of $\boldsymbol{w}$ orthogonal to $\boldsymbol{v}$



Example 2.6.8. Find the **distance** of the point $P = (7, 14)$ to the line $L : x - 2y = -6$.

The points $Q = (0, 3)$ and $R = (2, 4)$ are on the line L, and the vector $\boldsymbol{v} = \overline{QR} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ represents the direction of the line. Let $\boldsymbol{w} = \overline{QP} = \begin{pmatrix} 7 \\ 11 \end{pmatrix}$. The vector $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$ is perpendicular to the line and its **length** is the **distance** of the point p to the line L. We compute $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$ and its **length**:

$$Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w}) = \begin{pmatrix} 7 \\ 11 \end{pmatrix} - \frac{\begin{pmatrix} 7 \\ 11 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}}{\begin{pmatrix} 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 11 \end{pmatrix} - \frac{25}{5} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 11 \end{pmatrix} - \begin{pmatrix} 10 \\ 5 \end{pmatrix} = \begin{pmatrix} -3 \\ 6 \end{pmatrix}$$

$$\| \, Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w}) \, \| = \| \begin{pmatrix} -3 \\ 6 \end{pmatrix} \| = \sqrt{(-3)^2 + 6^2} = \sqrt{45} = 3\sqrt{5}.$$

Exercises

1. In each part find the **dot product** of the two given vectors $\boldsymbol{u}, \boldsymbol{v}$.

a) $\boldsymbol{u} = \begin{pmatrix} 2 \\ -5 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$.

b) $\boldsymbol{u} = \begin{pmatrix} 1 \\ -3 \\ 6 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 12 \\ 2 \\ -1 \end{pmatrix}$.

c) $\boldsymbol{u} = \begin{pmatrix} 2 \\ -3 \\ 4 \\ -1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} -2 \\ 4 \\ 1 \\ -6 \end{pmatrix}$.

d) $\boldsymbol{u} = \begin{pmatrix} 2 \\ -1 \\ 3 \\ -4 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 6 \\ 4 \\ -5 \\ -2 \end{pmatrix}$.

2. In each part find the **length** of the given **n-vector**.

a) $\begin{pmatrix} 3 \\ -3 \end{pmatrix}$.

b) $\begin{pmatrix} -3 \\ 12 \\ 4 \end{pmatrix}$.

c) $\begin{pmatrix} 1 \\ -2 \\ 4 \\ 2 \end{pmatrix}$.

d) $\begin{pmatrix} 4 \\ -7 \\ -4 \end{pmatrix}$.

3. For each vector in exercise 2 find a **unit vector** in the same direction.

4. In each part determine if the two vectors are **orthogonal**.

a) $\boldsymbol{u} = \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.

b) $\boldsymbol{u} = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} -6 \\ 8 \\ 10 \end{pmatrix}$.

c) $\boldsymbol{u} = \begin{pmatrix} 3 \\ 5 \\ 7 \\ 15 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}$.

d) $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} -4 \\ -3 \\ -2 \\ -1 \end{pmatrix}$.

In exercises 5 - 7 find the **distance** between all pairs from the given collection of **n-vectors**.

5. $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 7 \\ 5 \\ 1 \end{pmatrix}$.

6. $\boldsymbol{v}_1 = \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 3 \\ 4 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$.

7. $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \\ -2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 1 \\ -2 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 4 \\ 1 \\ 3 \\ -4 \end{pmatrix}$.

In exercises 8 and 9 find the **angle** between all pairs from the given collection of **n-vectors**.

8. $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}$.

9. $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

For each pair of vectors in exercises 10 - 13 verify the **Cauchy-Schwartz inequality** holds.

10. $\boldsymbol{x} = \begin{pmatrix} 2 \\ -3 \\ 1 \\ 4 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 3 \\ -2 \\ 1 \\ 1 \end{pmatrix}$

11. $\boldsymbol{x} = \begin{pmatrix} 3 \\ -4 \\ 12 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} -3 \\ 4 \\ 12 \end{pmatrix}$

12. $\boldsymbol{x} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 0 \\ -1 \\ -1 \\ 1 \end{pmatrix}$

13. $\boldsymbol{x} = \begin{pmatrix} 1 \\ -2 \\ 2 \\ -4 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 3 \\ -6 \\ 6 \\ -12 \end{pmatrix}$

For each pair of vectors in exercises 14-17 verify that the **triangle inequality** holds.

14. $\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 4 \\ 8 \\ -8 \end{pmatrix}$

15. $\boldsymbol{x} = \begin{pmatrix} 1 \\ -1 \\ 3 \\ 2 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} -6 \\ 4 \\ 2 \\ 5 \end{pmatrix}$

16. $\boldsymbol{x} = \begin{pmatrix} 4 \\ -6 \\ 8 \\ -2 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} -2 \\ 3 \\ -4 \\ 1 \end{pmatrix}$

17. $\boldsymbol{x} = \begin{pmatrix} -3 \\ 4 \\ 0 \\ 12 \end{pmatrix}, \begin{pmatrix} -4 \\ 3 \\ 12 \\ 0 \end{pmatrix}$

In exercises 18 - 21 for the given vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ compute a) $Proj_{\boldsymbol{v}}(\boldsymbol{w})$ and b) $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$. See **Method** (2.6.3).

18. $\boldsymbol{v} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} 5 \\ 1 \\ 3 \end{pmatrix}.$

19. $\boldsymbol{v} = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} -6 \\ 4 \\ 3 \end{pmatrix}.$

20. $\boldsymbol{v} = \begin{pmatrix} 1 \\ -2 \\ 4 \\ 2 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} 4 \\ 2 \\ 11 \\ 3 \end{pmatrix}$.

21. $\boldsymbol{v} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 3 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} 5 \\ 3 \\ 9 \\ 7 \end{pmatrix}$.

In exercises 22 and 23 find the length of the given point from the given line. See **Method** (2.6.4).

22. The point $P = (8, 7)$ and the line $L : 3x + 4y = 12$.

23. The point $P = (4, 8)$ and the line $L : y = \frac{1}{4}x + 2$.

24. Find all vectors in $\mathbb{R}^4$ which are **orthogonal** $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ where

$\boldsymbol{u}_1 = \begin{pmatrix} 2 \\ 3 \\ 1 \\ -2 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 0 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 3 \\ 5 \\ 1 \\ -6 \end{pmatrix}$.

See **Method** (2.6.1).

25. Find a **basis** for the **orthogonal complement** to $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ where

$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 2 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \end{pmatrix}$.

See **Method** (2.6.1).

26. For $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ the vectors in exercise 24, find all vectors $\boldsymbol{x} \in \mathbb{R}^4$ which satisfy

$$\boldsymbol{x} \cdot \boldsymbol{u}_1 = 2, \boldsymbol{x} \cdot \boldsymbol{u}_2 = -3, \boldsymbol{x} \cdot \boldsymbol{u}_3 = 9.$$

See **Method** (2.6.2).

27. For $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ the vectors in exercise 24, find all vectors $\boldsymbol{x} \in \mathbb{R}^4$ which satisfy

$$\boldsymbol{x} \cdot \boldsymbol{u}_1 = 1, \boldsymbol{x} \cdot \boldsymbol{u}_2 = 1, \boldsymbol{x} \cdot \boldsymbol{u}_3 = 1.$$

28. For $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ the vectors in exercise 25, find all vectors $\boldsymbol{x} \in \mathbb{R}^4$ which satisfy

$$\boldsymbol{x} \cdot \boldsymbol{v}_1 = 1, \boldsymbol{x} \cdot \boldsymbol{v}_2 = -1, \boldsymbol{x} \cdot \boldsymbol{v}_3 = 2.$$

See **Method** (2.6.2).

In exercises 29-33 answer true or false and give an explanation.

29. The **dot product** of any vector with the zero vector is zero.

30. If $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ then either $\boldsymbol{u} = \boldsymbol{0}_n$ of $\boldsymbol{v} = \boldsymbol{0}_n$.

31. If $\boldsymbol{u}, \boldsymbol{v}$ and $\boldsymbol{w}$ are in $\mathbb{R}^n$ and $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u} \cdot \boldsymbol{w}$ then $\boldsymbol{v} = \boldsymbol{u}$.

32. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are non-zero vectors and $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ then the **angle** between $\boldsymbol{u}$ and $\boldsymbol{v}$ is $\frac{\pi}{2}$.

33. The following holds for any three vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in $\mathbb{R}^n$:

$$(\boldsymbol{u} \cdot \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot (\boldsymbol{v} \cdot \boldsymbol{w}).$$

Challenge Exercises (Problems)

1. Suppose $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \in \mathbb{R}^n$ are non-zero vectors and $\boldsymbol{v}_1 \perp \boldsymbol{v}_2, \boldsymbol{v}_1 \perp \boldsymbol{v}_3, \boldsymbol{v}_2 \perp \boldsymbol{v}_3$. Prove that $\boldsymbol{v}_3 \notin Span(\boldsymbol{v}_1, \boldsymbol{v}_2)$.

2. Assume the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$of **3-vectors** is **linearly independent**.

a) Explain why $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = \mathbb{R}^3$.

b) Assume a vector $\boldsymbol{u} \neq \boldsymbol{0}_3$ satisfies $\boldsymbol{u} \cdot \boldsymbol{v}_2 = \boldsymbol{u} \cdot \boldsymbol{v}_3 = 0$. Prove that $\boldsymbol{u} \cdot \boldsymbol{v}_1 \neq 0$.

c) Prove that there are vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ such that

$\boldsymbol{u}_1 \cdot \boldsymbol{v}_1 = \boldsymbol{u}_2 \cdot \boldsymbol{v}_2 = \boldsymbol{u}_3 \cdot \boldsymbol{v}_3 = 1$ and

$\boldsymbol{u}_1 \cdot \boldsymbol{v}_2 = \boldsymbol{u}_1 \cdot \boldsymbol{v}_3 = \boldsymbol{u}_2 \cdot \boldsymbol{v}_1 = \boldsymbol{u}_2 \cdot v_3 = \boldsymbol{u}_3 \cdot \boldsymbol{v}_1 = \boldsymbol{u}_3 \cdot \boldsymbol{v}_2 = 0.$

3. Assume that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent** and **spans** $\mathbb{R}^3$. Assume the vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ satisfy

$\boldsymbol{v}_1 \cdot \boldsymbol{w}_1 = \boldsymbol{v}_2 \cdot \boldsymbol{w}_2 = \boldsymbol{v}_3 \cdot \boldsymbol{w}_3 = 1$ and

$\boldsymbol{v}_1 \cdot \boldsymbol{w}_1 = \boldsymbol{v}_1 \cdot \boldsymbol{w}_2 = \boldsymbol{v}_2 \cdot \boldsymbol{w}_1 = \boldsymbol{v}_2 \cdot \boldsymbol{w}_3 = \boldsymbol{v}_3 \cdot \boldsymbol{w}_1 = \boldsymbol{v}_3 \cdot \boldsymbol{w}_2 = 0.$
Prove that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$is **linearly independent** and **spans** $\mathbb{R}^3$.

4. a) Verify that the sequence $\left(\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix}\right) = (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent** and **spans** $\mathbb{R}^3$.

b) Find vectors $\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3$ such that $\boldsymbol{w}_1 \cdot \boldsymbol{z}_1 = \boldsymbol{w}_2 \cdot \boldsymbol{z}_2 = \boldsymbol{w}_3 \cdot \boldsymbol{z}_3 = 1$ and $\boldsymbol{w}_1 \cdot \boldsymbol{z}_2 = \boldsymbol{w}_1 \cdot \boldsymbol{z}_3 = \boldsymbol{w}_2 \cdot \boldsymbol{z}_1 = \boldsymbol{w}_2 \cdot \boldsymbol{z}_3 = \boldsymbol{w}_3 \cdot \boldsymbol{z}_1 = \boldsymbol{w}_3 \cdot \boldsymbol{z}_2 = 0.$

5. Use the converse to the **Pythagorean theorem** to prove that the given vertices span a right triangle.

$$\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} 2 \\ 1 \\ 4 \\ 3 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 13 \\ 6 \\ 5 \\ -3 \end{pmatrix}.$$

6. a) Prove that the following three points are the vertices of an **equilateral triangle**:

$$p_1 = \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}, p_2 = \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}, p_3 = \begin{pmatrix} 0 \\ 3 \\ 4 \end{pmatrix}.$$

b) Find an point p_4 such that p_1, p_2, p_3, p_4 are the vertices of a **regular tetrahedron**, i.e. such that all the faces are **equilateral triangles**.

c) Find the center of the **tetrahedron** of part b).

3. Prove the **law of cosines**: if a triangle has sides of length a, b, c and the angle opposite the side of length c is θ then $c^2 = a^2 + b^2 + 2ab\ cos\ \theta$. (Hint: Treat the sides of the triangle as vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}$ with $\| \boldsymbol{x} \| = a, \| \boldsymbol{y} \| = b$ and $\| \boldsymbol{x}+\boldsymbol{y} \| = c$.)

7. a) Verify the following identity for any two vectors $\boldsymbol{x}, \boldsymbol{y}$ in $\mathbb{R}^n$:

$$\| \boldsymbol{x}+\boldsymbol{y} \|^2 \ + \ \| \boldsymbol{x}-\boldsymbol{y} \|^2 \ = \ 2(\| \boldsymbol{x} \|^2 \ + \ \| \boldsymbol{y} \|^2)$$

b) Consider a parallelogram in the plane with sides of length a and b and diagonals of length d_1 and d_2. Prove that $d_1^2 + d_2^2 = 2(a^2 + b^2)$. (Hint: Make use of part a)).

Quiz Solutions

1. Compute the sum of the given vectors

a) $\begin{pmatrix} 1 \\ -3 \\ 2 \\ -5 \end{pmatrix} + \begin{pmatrix} -9 \\ 5 \\ 2 \\ -4 \end{pmatrix} = \begin{pmatrix} -8 \\ 2 \\ 4 \\ -9 \end{pmatrix}$

b) $\begin{pmatrix} 1 \\ -3 \\ 6 \end{pmatrix} + \begin{pmatrix} 4 \\ -2 \\ -2 \end{pmatrix} = \begin{pmatrix} 5 \\ -5 \\ 4 \end{pmatrix}.$

c) $\begin{pmatrix} 2 \\ -3 \\ 4 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ does not make sense since it is not possible to add a 4-vector and a 5-vector.

Not right, see **addition of $n-$vectors**.

2. Compute the scalar product $c\overline{v}$:

a) $\frac{1}{2}\overline{v} = \begin{pmatrix} -2 \\ 0 \\ 4 \\ 3 \end{pmatrix}$.

b) $-3\overline{v} = \begin{pmatrix} 3 \\ -6 \\ -15 \\ 6 \end{pmatrix}$.

c) $-1\overline{v} = \begin{pmatrix} 2 \\ -3 \\ 1 \\ -6 \\ 5 \end{pmatrix}$.

Not right, see **scalar multiplication of $n-$vectors**.

3. $Span\left(\begin{pmatrix} 5 \\ -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \\ 0 \\ 1 \end{pmatrix}\right)$.

Not right, see **Method** (1.2.4).

4. $\{\begin{pmatrix} -6 \\ 3 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} 7 \\ -3 \\ 1 \\ 0 \end{pmatrix} : t \in \mathbb{R}\}$.

Not right, see **Method** (1.2.4).

Chapter 3

Matrix Algebra

3.1. Introduction to Linear Transformations and Matrix Multiplication

In this section we define the notion of linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. We associate to each such transformation an $m \times n$ matrix called the **matrix** of the transformation.

Am I Ready for This Material

In order to follow this section and master its ideas and methods you need to be familiar with many previously defined concepts. **Those that are used extensively in this section are**:

an $n-$vector

equality of $n-$vectors

$n-$space, $\mathbb{R}^n$

adddition of $n-$vectors

scalar multiplication of an $n-$vector

negative of a vector

the zero vector

the standard basis for $\mathbb{R}^n$

linear combination of vectors

matrix

Quiz

Let $\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 2 \\ 1 \\ 3 \\ 2 \end{pmatrix}, \boldsymbol{u}_4 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}$. In each of the following, for the given **4-vector** $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$ compute the **linear combination** $c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + c_3\boldsymbol{u}_3 + c_4\boldsymbol{u}_4$.

1. $\begin{pmatrix} 4 \\ -1 \\ -1 \\ 0 \end{pmatrix}$

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

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

4. $\begin{pmatrix} 3 \\ 2 \\ -4 \\ 1 \end{pmatrix}$

Quiz Solutions

New Concepts

We begin this section with a review of the concept of a function (transformation, map) which should be familiar from calculus classes. We also introduce related concepts relevant to the functions which we will study in linear algebra, namely, linear transformations. These concepts are:

function, transformation, map

domain of a transformation

codomain of a transformation

image of an element under a transformation

range of a transformation

linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$

standard matrix of a linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$

product of an $m \times n$ matrix A and an $n-$vector $\boldsymbol{x}$

Theory (Why it Works)

The concept of a **function** is fundamental in mathematics. In calculus one studies how a function from $\mathbb{R}$ to $\mathbb{R}$ changes, with the emphasis on those functions which are *continuous* (where small changes in the input result in small changes in the output) and are *smooth* (don't abruptly change direction).

So far in linear algebra functions have played no explicit role. It is the purpose of the current section to begin to fill this gap. The functions that we will be concerned with in linear algebra will be ***linear***. Before we get to the definition, let's recall some concepts and terminology relevant to the study of functions.

First of all, what is a function? Intuitively, a function is a well defined rule that tells one in an unambiguous way how to assign an "output" for each "imput." We make this more precise.

Definition 3.1. A ***function, transformation, or map*** T from a set A to a set B is a rule which assigns to each member a of A a unique member b of B. This is denoted by $T : A \to B$ which is read, "T is a function (transformation, map) from A to B."

The set A is called the ***domain*** and the set B is the ***codomain***.

When T is a function (transformation, map) from the set A to B, as a metaphor, you can think of B as a target, where things from A are being shot and T is the mechanism that propels things from A to B.

Definition 3.2. When a function (transformation, map) T sends the element a of A to the element b of B we write $T(a) = b$ and say that T ***maps*** a to b and b is the ***image*** of a under T.

Let T be a **function** from A to B. Nothing in the definition requires that all the elements in (the "target") B actually get hit, or even very many, though something must since all the things in A are sent into B. We give a name to the collection of those elements of B which are **images** of elements from A :

Definition 3.3. Let $T : A \to B$ be a **function** from the set A to the set B. The set (collection) of those elements in B which are **images** of elements of A is called the ***range*** of T. We denote the range of T by $R(T)$. In set notation we have

$$R(T) := \{T(a) : a \in A\}.$$

In mathematics the terms **function, transformation and map** are used interchangeably and are synonyms. However, in different areas of mathematics one term gets used while in another area a different one may be more common. So, in calculus we typically use the term function. In abstract algebra, which deals with groups and rings, we more often use the map terminology. In linear algebra the most common usage is transformation, probably because the concept arose from generalizing functions which transformed the plane. The functions we will be interested in are called linear transformations, which we now will define:

Definition 3.4. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a ***linear transformation*** if

1) For all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n, T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$; and

2) For any $\boldsymbol{u} \in \mathbb{R}^n$ and scalar $c, T(c\boldsymbol{u}) = cT(\boldsymbol{u})$.

We will refer to the first condition in **Definition** (3.4) as the "additive property" of a linear transformation and the second condition as the "scalar property".

Example 3.1.1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x \\ 3y \end{pmatrix}$. Then T is a linear transformation. To see that this is so we have to check that the additive and scalar properties hold.

1) The additive property holds: Let $\boldsymbol{u} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$. Then $\boldsymbol{u}+\boldsymbol{v} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}$.

$T(\boldsymbol{u} + \boldsymbol{v}) = T\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} = \begin{pmatrix} 2(x_1 + x_2) \\ 3(y_1 + y_2) \end{pmatrix}$ by the definition of T;

$= \begin{pmatrix} 2x_1 + 2x_2 \\ 3y_1 + 3y_3 \end{pmatrix}$ by the distributive property for real multiplication;

$= \begin{pmatrix} 2x_1 \\ 3y_1 \end{pmatrix} + \begin{pmatrix} 2x_2 \\ 3y_2 \end{pmatrix}$ by the definition of **vector addition**;

$= T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = T(\boldsymbol{u}) + T(\boldsymbol{v})$ by the definition of T.

2) The scalar property holds: Now let $\boldsymbol{w} = \begin{pmatrix} x \\ y \end{pmatrix}$ and c be a scalar. We need to show that $T(c\boldsymbol{w}) = cT(\boldsymbol{w})$. Now $c\boldsymbol{w} = c\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} cx \\ cy \end{pmatrix}$. Then $T(c\boldsymbol{w}) = T\begin{pmatrix} c\boldsymbol{x} \\ c\boldsymbol{y} \end{pmatrix} = \begin{pmatrix} 2(cx) \\ 2(cy) \end{pmatrix}$ by the definition of T;

$= \begin{pmatrix} c(2x) \\ c(3y) \end{pmatrix}$ by the associativity and commutativity of real multiplication. In turn, this is equal to

$= c\begin{pmatrix} 2x \\ 3y \end{pmatrix} = cT(\boldsymbol{w})$ by the definition of **scalar multiplication** and the definition of T.

The next example is more abstract.

Example 3.1.2. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ be vectors in $\mathbb{R}^2$. Define $T : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + x_3\boldsymbol{v}_3 \tag{3.1}$$

Then T is a linear transformation.

We show that 1) and 2) of **Definition** (3.4) hold.

1) Let $\boldsymbol{x}_1 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$. Then $\boldsymbol{x} + \boldsymbol{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix}$.

By the definition of T,

$$T(\boldsymbol{x} + \boldsymbol{y}) = T \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} = (x_1 + y_1)\boldsymbol{v}_1 + (x_2 + y_2)\boldsymbol{v}_2 + (x_3 + y_3)\boldsymbol{v}_3 =$$

$$[x_1\boldsymbol{v}_1 + y_1\boldsymbol{v}_1] + [x_2\boldsymbol{v}_2 + y_2\boldsymbol{v}_2] + [x_3\boldsymbol{v}_3 + y_3\boldsymbol{v}_3]$$

by the **distributive property of scalar multiplication**;

$$= [x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + x_3\boldsymbol{v}_3] + [y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 + y_3\boldsymbol{v}_3]$$

by the **associativity and commutativity of vector addition**;

$$= T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + T \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = T(\boldsymbol{x}) + T(\boldsymbol{y})$$

by the definition of T.

2) . Now let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ be a **3-vector** and c a scalar. Then $c\boldsymbol{x} = \begin{pmatrix} cx_1 \\ cx_2 \\ cx_3 \end{pmatrix}$. Therefore

$$T(c\boldsymbol{x}) = T \begin{pmatrix} cx_1 \\ cx_2 \\ cx_3 \end{pmatrix} = (cx_1)\boldsymbol{v}_1 + (cx_2)\boldsymbol{v}_2 + (cx_3)\boldsymbol{v}_3 = c(x_1\boldsymbol{v}_1) + c(x_2\boldsymbol{v}_2) + c(x_3\boldsymbol{v}_3),$$

the first equality follows by the definition of T and the second by part 7) of **Theorem** (2.2.1). In turn, this is equal to

$$c[x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + x_3\boldsymbol{v}_3]$$

by properties of **addition** and **scalar multiplication** as demonstrated in **Theorem** (2.2.1).

$$= cT\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = cT(\boldsymbol{x})$$

by the definition of T.

When we define a function in this way it always gives rise to a **linear transformation**, which is the subject of our first theorem. We omit the proof since it is exactly as in **Example** (3.1.2).

Theorem 3.1.1. *Let $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a sequence of* **m-vectors** *. Define a function $T : \mathbb{R}^n \to \mathbb{R}^m$ as follows:*

$$T\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n \tag{3.2}$$

Then T is a **linear transformation***.*

As we will see shortly, it is possible to give a characterization of all **linear transformations** from $\mathbb{R}^n$ to $\mathbb{R}^m$. In fact, they all arise as in Theorem (3.1.1) above. Before showing this we prove a necessary result that tells us how a **linear transformation** acts on a **linear combination** of vectors.

Theorem 3.1.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a* **linear transformation***. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ be vectors in $\mathbb{R}^n$ and $c_1, c_2, \ldots, c_k$ be scalars. Then*

$$T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \cdots + c_3T(\boldsymbol{v}_k) \tag{3.3}$$

Proof. When $k = 1$ this is just the second property of a **linear transformation** and there is nothing to prove.

When $k = 2$ we have: $T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = T(c_1\boldsymbol{v}_1) + T(c_2\boldsymbol{v}_2) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2)$, the first equality by the **additive property of linear transformations** and the second equality by the **scalar property of linear transformations**.

The general proof is by the **principle of mathematical induction** on k. All the necessary arguments are present when we pass from the case $k = 2$ to the case $k = 3$ and so we illustrate the proof in this case and leave the general case to the student as a problem.

By the **associativity property for vector addition**, $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = [c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2] + c_3\boldsymbol{v}_3$ a sum of the two vectors $[c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2]$ and $c_3\boldsymbol{v}_3$.

By the **additive property of linear transformations** we have

$$T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3) = T([c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2] + c_3\boldsymbol{v}_3) = T([c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2]) + T(c_3\boldsymbol{v}_3).$$

By another application of the **additive property of linear transformations** (to the first term) we get

$$= T(c_1\boldsymbol{v}_1) + T(c_2\boldsymbol{v}_3) + T(c_3\boldsymbol{v}_3) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + c_3T(\boldsymbol{v}_3),$$

the last equality by application of the **scalar property of linear transformations**.

□

Making use of **Theorem** (3.1.2) we can now classify the **linear tranformations** from $\mathbb{R}^n \to \mathbb{R}^m$:

Theorem 3.1.3. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. Let $\mathbf{e}_i^n$ be the i^{th} **standard basis vector** of $\mathbb{R}^n$ and set $\mathbf{v}_i = T(\mathbf{e}_i^n)$. Then*

$$T\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n.$$

Proof. The proofs for arbitrary n and for the particular case $n = 3$ are exactly the same and therefore we do this specific case. First note that $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3$.

Then by **Theorem** (3.1.2) we have

$$\begin{aligned} T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= T(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3) \\ &= x_1T(\boldsymbol{e}_1) + x_2T(\boldsymbol{e}_2) + x_3T(\boldsymbol{e}_3) \\ &= x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + x_3\boldsymbol{v}_3 \end{aligned}$$

since $T(\boldsymbol{e}_i) = \boldsymbol{v}_i$ for $i = 1, 2, 3$. □

Theorem (3.1.3) says that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** and we know the **images** of the **the standard basis vectors** then in principle we know the effect of T on all vectors. When we know the **images** of the **the standard basis vectors** we can encode this information in a matrix:

Definition 3.5. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. Set $\boldsymbol{v}_i = T(\boldsymbol{e}_i^n), 1 \leq i \leq n$ where $\boldsymbol{e}_i^n$ is the i^{th} **the standard basis vector** of $\mathbb{R}^n$. Then the $m \times n$ martrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_n)$ is called the ***standard matrix of*** T.

We make use of the **standard matrix** to define the product $A\boldsymbol{x}$ of an $m \times n$, A matrix and an **n-vector**, $\boldsymbol{x}$.

Definition 3.6. Let A be an $m \times n$ matrix with columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ and let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ be an **n-vector**. Then the ***product*** $A\boldsymbol{x}$ is defined and is equal to

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \ldots x_n\boldsymbol{v}_n.$$

Remark 3.1. i) If A is an $m \times n$ matrix and we define $T_A : \mathbb{R}^n \to \mathbb{R}^m$ by $T_A\boldsymbol{x} = A\boldsymbol{x}$ then T_A is a **linear transformation** by **Theorem** (3.1.1). We refer to a **linear transformation** that arises this way as a ***matrix transformation***. Since A can be an arbitrary $m \times n$ matrix, this implies that every $m \times n$ matrix is the **standard matrix** of some **linear transformation** from $\mathbb{R}^n$ to $\mathbb{R}^m$.

ii) **Theorem** (3.1.3) says that every **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **matrix transformation**.

Example 3.1.3. Show that the transformation given below is a **linear transformation** and find its **standard matrix**:

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x-2y+z \\ 2x-3y+z \\ 4x-7y+3z \end{pmatrix}$$

$$\begin{pmatrix} x-2y+z \\ 2x-3y+z \\ 4x-7y+3z \end{pmatrix} = \begin{pmatrix} x \\ 2x \\ 4x \end{pmatrix} + \begin{pmatrix} -2y \\ -3y \\ -7y \end{pmatrix} + \begin{pmatrix} z \\ z \\ 3z \end{pmatrix} =$$

$$x\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + y\begin{pmatrix} -2 \\ -3 \\ -7 \end{pmatrix} + z\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ 2 & -3 & 1 \\ 4 & -7 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Thus T is a **matrix transformation** with **standard matrix** $\begin{pmatrix} 1 & -2 & 1 \\ 2 & -3 & 1 \\ 4 & -7 & 3 \end{pmatrix}$.

We next prove a two results making explicit properties of **linear transformations** which will be handy later on.

Theorem 3.1.4. *Assume $T : \mathbb{R}^n \to \mathbb{R}^m$ is a* **linear transformation**. *Then the following hold:*

1) $T(\mathbf{0}_n) = \mathbf{0}_m$.

2) If $\boldsymbol{v}$ is an **n-vector** *then $T(-\boldsymbol{v}) = -T(\boldsymbol{v})$.*

3) The **range** *of T is a* **subspace** *of $\mathbb{R}^m$.*

Proof. 1) Set $\boldsymbol{u} = T(\mathbf{0}_n)$ (note that this is an **m-vector** and so $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}_m$). Since for any scalar $c, c\mathbf{0}_n = \mathbf{0}_n$, in particular, $2\mathbf{0}_n = \mathbf{0}_n$. It then follows that $T(2\mathbf{0}_n) = T(\mathbf{0}_n)$. Now by the **scalar property** we get

$$2T(\mathbf{0}_n) = T(\mathbf{0}_n)$$

$$2\boldsymbol{u} = \boldsymbol{u} \tag{3.4}$$

Adding $-\boldsymbol{u}$ to both sides of Equation (3.4) we obtain $\boldsymbol{u} = \mathbf{0}_m$ as required.

2) Set $T(\boldsymbol{v}) = \boldsymbol{u}$. Now $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}_n$. Using part 1) we get

$$T(\boldsymbol{v} + (-\boldsymbol{v})) = T(\mathbf{0}_n) = \mathbf{0}_m.$$

By the **additive property of linear transformations** we have

$$T(\boldsymbol{v}) + T(-\boldsymbol{v}) = \mathbf{0}_m$$

$$\boldsymbol{u} + T(-\boldsymbol{v}) = \mathbf{0}_m. \tag{3.5}$$

Adding $-\boldsymbol{u} = -T(\boldsymbol{v})$ to both sides of Equation (3.5) we obtain

$$T(-\boldsymbol{v}) = -\boldsymbol{u} = -T(\boldsymbol{v}).$$

3) By part 1) we know that $\mathbf{0}_m = T(\mathbf{0}_n)$ is in $R(T)$ and so it is non-empty. We need to show that it is **closed under addition and multiplication**.

Closed under addition: Assume $\boldsymbol{u}_1, \boldsymbol{u}_2$ are **m-vectors** in $R(T)$. Then there are vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ from $\mathbb{R}^n$ such that $T(\boldsymbol{v}_1) = \boldsymbol{u}_1$ and $T(\boldsymbol{v}_2) = \boldsymbol{u}_2$. By the **additive property** of a **linear transformation**, $T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = T(\boldsymbol{v}_1) + T(\boldsymbol{v}_2) = \boldsymbol{u}_1 + \boldsymbol{u}_2$. Thus, $\boldsymbol{u}_1 + \boldsymbol{u}_2$ is in $R(T)$ and $R(T)$ is, indeed, closed under addition.

Closed under scalar multiplication: Assume $\boldsymbol{u}$ is an **m-vector** and $\boldsymbol{u}$ is in the **range** of T and that c is a scalar. Since $\boldsymbol{u}$ is in $R(T)$ there is an **n-vector**, $\boldsymbol{v}$, such that $T(\boldsymbol{v}) = \boldsymbol{u}$. By the **scalar property** of a **linear transformation**, $T(c\boldsymbol{v}) = cT(\boldsymbol{v}) = c\boldsymbol{u}$. Thus, $c\boldsymbol{u}$ is in $R(T)$ and $R(T)$ is closed under scalar multiplication. Therefore $R(T)$ is a **subspace** . □

We close this section with a couple of examples demonstrating that some common functions of $\mathbb{R}^n$ are **linear transformations**. In our first example we show that rotation of the plane is a matrix transformation of $\mathbb{R}^2$ to $\mathbb{R}^2$.

Example 3.1.4. Rotation about the origin by an angle of θ radians is a **matrix transformation** with **standard matrix** equal to

$$\begin{pmatrix} cos\ \theta & -sin\ \theta \\ sin\ \theta & cos\ \theta \end{pmatrix}$$

Denote by R_θ the function of the plane ($\mathbb{R}^2$) which rotates a vector through an angle θ in the counterclockwise direction. Consider a typical vector $\begin{pmatrix} x \\ y \end{pmatrix} \neq \mathbf{0}_2$. Set $r = \sqrt{x^2 + y^2} = \| \begin{pmatrix} x \\ y \end{pmatrix} \|$ and $\phi = \arctan \frac{y}{x}$. When $\boldsymbol{x}$ is in the first or fourth quadrants then $\boldsymbol{x} = r\begin{pmatrix}\cos\phi\\ \sin\phi\end{pmatrix}$. On the other hand, if $\boldsymbol{x}$ is in the second or third quadrants then $\boldsymbol{x} = (-r)\begin{pmatrix}\cos\phi\\ \sin\phi\end{pmatrix}$.

We first treat the case that $\boldsymbol{x}$ is in the first or fourth quadrants. As noted above, $x = r\ \cos\phi, y = r\sin\phi$. When we rotate in the counterclockwise direction by an angle θ the vector $r\begin{pmatrix}\cos\phi\\ \sin\phi\end{pmatrix}$ is taken to the vector $r\begin{pmatrix}\cos\phi+\theta\\ \sin\phi+\theta\end{pmatrix}$.

Recall the **formulas** for $\cos(\phi+\theta)$ and $\sin(\phi+\theta)$:

$$\cos(\phi+\theta) = \cos\phi\cos\theta - \sin\phi\sin\theta$$

$$\sin(\phi+\theta) = \cos\phi\sin\theta + \sin\phi\sin\theta.$$

Then

$$x' = r\cos\phi+\theta = r(\cos\phi\cos\theta - \sin\phi\sin\theta) =$$

$$(r\cos\phi)\cos\theta + (r\sin\phi)(-\sin\theta) = x\cos\theta - y\sin\theta$$

$$y' = r\sin\phi+\theta = r(\cos\phi\sin\theta + \sin\phi\sin\theta) =$$

$$(r\cos\phi)(\sin\theta) + (r\sin\phi)\cos\theta = x\sin\theta + y\cos\theta.$$

It then follows that

$$R_\theta\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}x cos\ \theta - y sin\ \theta\\ x sin\ \theta + y cos\ \theta\end{pmatrix} =$$

$$\begin{pmatrix}x\ cos\ \theta\\ x\ sin\ \theta\end{pmatrix} + \begin{pmatrix}-y\ sin\ \theta\\ y\ cos\ \theta\end{pmatrix} =$$

$$x\begin{pmatrix}cos\ \theta\\ sin\ \theta\end{pmatrix} + y\begin{pmatrix}-sin\ \theta\\ cos\ \theta\end{pmatrix} = \begin{pmatrix}cos\ \theta & -sin\ \theta\\ sin\ \theta & cos\ \theta\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}.$$

When $\boldsymbol{v} = \begin{pmatrix}x\\ y\end{pmatrix}$ is in the second or third quadrants then we can write $\boldsymbol{v} = (-r)\begin{pmatrix}cos\ \phi\\ sin\ \phi\end{pmatrix}$ where $r = \sqrt{x^2+y^2}$ and $\phi = arctan\ \frac{y}{x}$. In this case we get $R_\theta(\boldsymbol{v}) = \begin{pmatrix}cos\ \theta & -sin\ \theta\\ sin\ \theta & cos\ \theta\end{pmatrix}\boldsymbol{v}$ as well.

Example 3.1.5. Recall in Section (2.6), if $\boldsymbol{v}, \boldsymbol{w}$ are **n-vectors** with $\boldsymbol{v} \neq \boldsymbol{0}_n$ we defined $Proj_{\boldsymbol{v}}(\boldsymbol{w})$, the **projection of w onto v**, and $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$, the **projection of w orthogonal to v** We claim these are **linear transformations** . We will demonstrate this for $Proj_{\boldsymbol{v}}(\boldsymbol{w})$. That $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{w})$ is a **linear transformation** we leave as a **challenge exercise**.

Theorem 3.1.5. *Let $\boldsymbol{v}$ be a non-zero* **n-vector**. *Then $Proj_{\boldsymbol{v}} : \mathbb{R}^n \to \mathbb{R}^n$ is a* **linear transformation**.

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and c be a scalar. We must show

1) $Proj_{\boldsymbol{v}}(\boldsymbol{x}+\boldsymbol{y}) = Proj_{\boldsymbol{v}}(\boldsymbol{x}) + Proj_{\boldsymbol{v}}(\boldsymbol{y})$; and

2) $Proj_{\boldsymbol{v}}(c\boldsymbol{x}) = cProj_{\boldsymbol{v}}(\boldsymbol{x})$.

1)
$$Proj_{\boldsymbol{v}}(\boldsymbol{x}+\boldsymbol{y}) = \frac{(\boldsymbol{x}+\boldsymbol{y})\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} =$$

$$\frac{\boldsymbol{x}\cdot\boldsymbol{v}+\boldsymbol{y}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} = (\frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}} + \frac{\boldsymbol{y}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}})\boldsymbol{v} =$$

$$\frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} + \frac{\boldsymbol{y}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} = Proj_{\boldsymbol{v}}(\boldsymbol{x}) + Proj_{\boldsymbol{v}}(\boldsymbol{y}).$$

2)
$$Proj_{\boldsymbol{v}}(c\boldsymbol{x}) = \frac{(c\boldsymbol{x})\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} =$$

$$\frac{c(\boldsymbol{x}\cdot\boldsymbol{v})}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} = (c\frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}})\boldsymbol{v} = c(\frac{\boldsymbol{x}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}) = cProj_{\boldsymbol{v}}(\boldsymbol{x}).$$

□

Example 3.1.6. Let $\boldsymbol{v} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$. Determine the **standard matrix** of $Proj_{\boldsymbol{v}}$ and $Proj_{\boldsymbol{v}^\perp}$.

$$Proj_{\boldsymbol{v}}(\boldsymbol{e}_1) = \frac{\boldsymbol{e}_1\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v} = \frac{1}{9}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{pmatrix},$$

$$Proj_{\boldsymbol{v}}(\boldsymbol{e}_2) = \frac{\boldsymbol{e}_2 \cdot \boldsymbol{v}}{9}\boldsymbol{v} = \frac{2}{9}\begin{pmatrix}1\\2\\2\end{pmatrix} = \begin{pmatrix}\frac{2}{9}\\\frac{4}{9}\\\frac{4}{9}\end{pmatrix},$$

$$Proj_{\boldsymbol{v}}(\boldsymbol{e}_3) = \frac{\boldsymbol{e}_3 \cdot \boldsymbol{v}}{9}\boldsymbol{v} = \frac{2}{9}\begin{pmatrix}1\\2\\2\end{pmatrix} = \begin{pmatrix}\frac{2}{9}\\\frac{4}{9}\\\frac{4}{9}\end{pmatrix}.$$

The **standard matrix** of $Proj_{\boldsymbol{v}}$ is $\begin{pmatrix}\frac{1}{9} & \frac{2}{9} & \frac{2}{9}\\\frac{2}{9} & \frac{4}{9} & \frac{4}{9}\\\frac{2}{9} & \frac{4}{9} & \frac{4}{9}\end{pmatrix}$.

$$Proj_{\boldsymbol{v}^\perp}(\boldsymbol{e}_1) = \boldsymbol{e}_1 - Proj_{\boldsymbol{v}}(\boldsymbol{e}_1) = \begin{pmatrix}1\\0\\0\end{pmatrix} - \begin{pmatrix}\frac{1}{9}\\\frac{2}{9}\\\frac{2}{9}\end{pmatrix} = \begin{pmatrix}\frac{8}{9}\\-\frac{2}{9}\\-\frac{2}{9}\end{pmatrix},$$

$$Proj_{\boldsymbol{v}^\perp}(\boldsymbol{e}_2) = \boldsymbol{e}_2 - Proj_{\boldsymbol{v}}(\boldsymbol{e}_2) = \begin{pmatrix}0\\1\\0\end{pmatrix} - \begin{pmatrix}\frac{2}{9}\\\frac{4}{9}\\\frac{4}{9}\end{pmatrix} = \begin{pmatrix}-\frac{2}{9}\\\frac{5}{9}\\-\frac{4}{9}\end{pmatrix},$$

$$Proj_{\boldsymbol{v}^\perp}(\boldsymbol{e}_3) = \boldsymbol{e}_3 - Proj_{\boldsymbol{v}}(\boldsymbol{e}_3) = \begin{pmatrix}0\\0\\1\end{pmatrix} - \begin{pmatrix}\frac{2}{9}\\\frac{4}{9}\\\frac{4}{9}\end{pmatrix} = \begin{pmatrix}-\frac{2}{9}\\-\frac{4}{9}\\\frac{5}{9}\end{pmatrix}.$$

The **standard matrix** of $Proj_{\boldsymbol{v}^\perp}$ is $\begin{pmatrix}\frac{8}{9} & -\frac{2}{9} & -\frac{2}{9}\\-\frac{2}{9} & \frac{5}{9} & -\frac{4}{9}\\-\frac{2}{9} & -\frac{4}{9} & \frac{5}{9}\end{pmatrix}$

What You Can Now Do

1. Given a **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by its **images** on the **standard basis vectors** $\boldsymbol{e}_1^n, \boldsymbol{e}_2^n, \dots, \boldsymbol{e}_n^n$ determine the **standard matrix** of T.

2. Given a **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ with **images** given by explicit formulas determine the **standard matrix** of T.

3. Given a **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by its **images** on the **standard basis vectors** $\boldsymbol{e}_1^n, \boldsymbol{e}_2^n, \dots, \boldsymbol{e}_n^n$ and an **n-vector**, $\boldsymbol{x}$, find the image of $\boldsymbol{x}$.

4. Given an $m \times n$ matrix A and an **n-vector**, $\boldsymbol{x}$, compute the **matrix product** $A\boldsymbol{x}$.

Method (How to do it)

Method 3.1.1. Given a **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by its **images** on the **standard basis vectors** $\boldsymbol{e}_1^n, \boldsymbol{e}_2^n, \dots, \boldsymbol{e}_n^n$ determine the **standard matrix** of T.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** and $T(\boldsymbol{e}_i^n) = \boldsymbol{v}_i$ then the **standard matrix** of T is the $m \times n$ matrix $(\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \dots \ \boldsymbol{v}_n)$.

Example 3.1.7. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a **linear transformation** and assume that $T\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\3\\-4\end{pmatrix}, T\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}2\\5\\-7\end{pmatrix}$. The **standard matrix** of T is $\begin{pmatrix}1 & 2\\3 & 5\\-4 & -7\end{pmatrix}$.

Example 3.1.8. Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be a **linear transformation** and assume that $T\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, T\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}1\\2\\4\\8\end{pmatrix}, T\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}1\\3\\9\\27\end{pmatrix}$. The **standard matrix** of T is $\begin{pmatrix}1 & 1 & 1\\1 & 2 & 3\\1 & 4 & 9\\1 & 8 & 27\end{pmatrix}$.

Method 3.1.2. Given a **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^m$ with **images** given by explicit formulas determine the **standard matrix** of T.

Use the formulas to compute the **images** of the **standard basis vectors**: $T(\boldsymbol{e}_1^n), T(\boldsymbol{e}_2^n), \dots, T(\boldsymbol{e}_n^n)$. The **standard matrix** of T is $(T(\boldsymbol{e}_1^n) \ T(\boldsymbol{e}_2^n) \ \dots \ T(\boldsymbol{e}_n^n))$.

Example 3.1.9. Assume that $T : \mathbb{R}^3 \to \mathbb{R}^2$ is a given by

$$T\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}x + 2y - 4z\\-3x + 5y - z\end{pmatrix}$$

We can use these formulas to compute:

$$T\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}1\\-3\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}2\\5\end{pmatrix},T\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}-4\\-1\end{pmatrix}.$$

The **standard matrix** of T is $\begin{pmatrix}1&2&-4\\-3&5&-1\end{pmatrix}$

Example 3.1.10. Assume that $T:\mathbb{R}^3\to\mathbb{R}^3$ is given by

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}-2x+5y-3z\\-x-2y+3z\\4x-5y+z\end{pmatrix}$$

Then

$$T\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}-2\\-1\\4\end{pmatrix},T\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}5\\-2\\-5\end{pmatrix},T\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}-3\\3\\1\end{pmatrix}.$$

The **standard matrix** of T is $\begin{pmatrix}-2&5&-3\\-1&-2&3\\4&-5&1\end{pmatrix}$

Method 3.1.3. Given a **linear transformation** $T:\mathbb{R}^n\to\mathbb{R}^m$ defined by its **images** on the **standard basis vectors** $\boldsymbol{e}_1^n,\boldsymbol{e}_2^n,\dots,\boldsymbol{e}_n^n$ and an **n-vector**, $\boldsymbol{x}$, find the image of $\boldsymbol{x}$.

If the **image** of $\boldsymbol{e}_i^n$ is $T(\boldsymbol{e}_i^n)=\boldsymbol{a}_i$ and $\boldsymbol{x}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}$ then

$$T(\boldsymbol{x})=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\cdots+x_n\boldsymbol{a}_n.$$

Example 3.1.11. Let T be the **linear transformation** of **Example** (3.1.7) and $\boldsymbol{x}=\begin{pmatrix}-2\\4\end{pmatrix}$. Find $T(\boldsymbol{x})$.

$$T(\boldsymbol{x})=(-2)\begin{pmatrix}1\\3\\-4\end{pmatrix}+(4)\begin{pmatrix}2\\5\\-7\end{pmatrix}=\begin{pmatrix}6\\14\\-20\end{pmatrix}.$$

Example 3.1.12. Let T be the **linear transformation** of **Example** (3.1.8) and $\boldsymbol{x} = \begin{pmatrix} 5 \\ -3 \\ 1 \end{pmatrix}$. Find $T(\boldsymbol{x})$.

$$T(\boldsymbol{x}) = (5)\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = (-3)\begin{pmatrix} 1 \\ 2 \\ 4 \\ 8 \end{pmatrix} + (1)\begin{pmatrix} 1 \\ 3 \\ 9 \\ 27 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 2 \\ 8 \end{pmatrix}$$

Method 3.1.4. Given an $m \times n$ matrix A and an **n-vector**, $\boldsymbol{x}$, compute the **matrix product** $A\boldsymbol{x}$.

If the columns of A are $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ and $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ then the **matrix product** $A\boldsymbol{x}$ is equal to $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + a_n\boldsymbol{x}_n$.

Example 3.1.13. Compute the **matrix product** $\begin{pmatrix} 1 & 4 & -5 \\ -3 & -3 & 6 \\ 2 & 1 & -3 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$.

$$\begin{pmatrix} 1 & 4 & -5 \\ -3 & -3 & 6 \\ 2 & 1 & -3 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} = (2)\begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} + (2)\begin{pmatrix} 4 \\ -3 \\ 1 \end{pmatrix} + (1)\begin{pmatrix} -5 \\ 6 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ -6 \\ 3 \end{pmatrix}$$

Example 3.1.14. Compute the **matrix product** $\begin{pmatrix} 5 & 2 \\ 3 & 1 \\ -3 & -2 \\ 6 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

$$\begin{pmatrix} 5 & 2 \\ 3 & 1 \\ -3 & -2 \\ 6 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix} = (1)\begin{pmatrix} 5 \\ 3 \\ -3 \\ 6 \end{pmatrix} + (-2)\begin{pmatrix} 2 \\ 1 \\ -2 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

Exercises

In exercises 1-4 compute the **standard matrix** of the given **linear transformation** T. See **Method** (3.1.1).

1. $T\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}4\\-2\\1\\0\end{pmatrix}, T\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}4\\2\\5\\-2\end{pmatrix}$

2. $T\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\-2\\4\\-8\end{pmatrix}, T\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}-1\\1\\-1\\1\end{pmatrix}, T\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}1\\-3\\9\\-27\end{pmatrix}.$

3. $T\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\1\\0\end{pmatrix}, T\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\end{pmatrix}, T\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}1\\0\\0\end{pmatrix}.$

4. $T\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\2\\-3\end{pmatrix}, T\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}2\\5\\-7\end{pmatrix}, T\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}-3\\-1\\4\end{pmatrix}.$

In exercises 5 - 7 for the given matrix A and pair of vectors $\boldsymbol{v}, \boldsymbol{w}$, compute the **matrix products** $A\boldsymbol{v}$, $A\boldsymbol{w}$ and $A(\boldsymbol{v}+\boldsymbol{w})$ and verify the equality $A(\boldsymbol{v}+\boldsymbol{w}) = A\boldsymbol{v}+A\boldsymbol{w}$. See **Method** (3.1.4).

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

$\boldsymbol{v} = \begin{pmatrix}1\\1\\2\end{pmatrix}, \boldsymbol{w} = \begin{pmatrix}2\\-1\\1\end{pmatrix}$

6. $A = \begin{pmatrix}1 & -1 & 0 & 3\\0 & 2 & -4 & 1\end{pmatrix}, \boldsymbol{v} = \begin{pmatrix}1\\-1\\0\\-1\end{pmatrix}, \boldsymbol{w} = \begin{pmatrix}4\\1\\3\\-1\end{pmatrix}$

7. $A = \begin{pmatrix}2 & -1 & -5 & -2 & 2\\-3 & 4 & 5 & 0 & 0\\-1 & -1 & 4 & 1 & -1\end{pmatrix}, \boldsymbol{v} = \begin{pmatrix}3\\1\\1\\0\\0\end{pmatrix}, \boldsymbol{w} = \begin{pmatrix}0\\0\\0\\1\\1\end{pmatrix}$

In exercises 8 - 10 for the given matrix A, vector $\boldsymbol{v}$ and scalar c, compute **matrix products** $A\boldsymbol{v}$, $A(c\boldsymbol{v})$ and $c(A\boldsymbol{v})$ and verify that $A(c\boldsymbol{v}) = c(A\boldsymbol{v})$. See **Method** (3.1.4).

8. $A = \begin{pmatrix} 3 & -2 & -1 \\ 0 & -4 & 4 \\ 2 & -3 & 1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, c = 2$

9. $A = \begin{pmatrix} 2 & -1 & -3 & 5 \\ 0 & 3 & -1 & 2 \\ 4 & 0 & 3 & -1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 2 \\ -1 \\ 1 \\ -1 \end{pmatrix}, c = -1$

10. $A = \begin{pmatrix} -3 & -2 & -1 & 1 \\ 4 & 3 & 1 & 1 \\ 6 & 2 & 1 & -2 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 1 \\ -1 \\ 2 \\ 3 \end{pmatrix}, c = 4$

In exercises11 - 14 compute the **standard matrix** of the given **linear transformation** T. See **Method** (3.1.2).

11. $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - 3y \\ 3x + 5y \\ x + 2y \\ -x - y \end{pmatrix}.$

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

13. $T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_2 + x_3 \\ x_3 + x_4 \\ x_1 + x_4 \\ x_1 + x_2 \end{pmatrix}$

14. $T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -x + y + z \\ x - y + z \\ x + y - z \\ 2x + 2y + 2z \end{pmatrix}$

In exercises 15 - 18 find the **image**, $T(\boldsymbol{x})$, for given **linear transformation** T and vector $\boldsymbol{x}$. See **Method** (3.1.3).

15. T is the transformation of Exercise 1 and $\boldsymbol{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

16. T is the transformation of Exercise 2 and $\boldsymbol{x} = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}$

17. T is the transformation of Exercise 3 and $\boldsymbol{x} = \begin{pmatrix} 5 \\ -7 \\ 3 \end{pmatrix}.$

18. T is the transformation of Exercise 4 and $\boldsymbol{x} = \begin{pmatrix} 13 \\ -5 \\ 1 \end{pmatrix}$.

In exercises 19 - 22 compute the **matrix product** $A\boldsymbol{v}$ for the given matrix A and vector $\boldsymbol{v}$. See **Method** (3.1.4).

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

a) $\boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ b) $\boldsymbol{v} = \begin{pmatrix} 3 \\ 3 \\ -2 \end{pmatrix}$ c) $\boldsymbol{v} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ d) $\boldsymbol{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ e) $\boldsymbol{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix}$

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

a) $\boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ b) $\boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ c) $\boldsymbol{v} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$ d) $\boldsymbol{v} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$

21. $A = \begin{pmatrix} 2 & -1 & -3 \\ 0 & -3 & 3 \\ -3 & -1 & 7 \end{pmatrix}$

a) $\boldsymbol{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ b) $\boldsymbol{v} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ c) $\boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ d) $\boldsymbol{v} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$

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

a) $\boldsymbol{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ b) $\boldsymbol{v} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix}$ c) $\boldsymbol{v} = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 3 \end{pmatrix}$ d) $\boldsymbol{v} = \begin{pmatrix} -1 \\ 2 \\ 3 \\ -2 \end{pmatrix}$

In exercises 23-27 answer true or false and give an explanation.

23. Every **function** $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation**.

24. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function then $T(\mathbf{0}_n) = \mathbf{0}_m$.

25. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** then $T(\mathbf{0}_n) = \mathbf{0}_m$.

26. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function then $\mathbf{0}_n$ is in the **range** of T.

27. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** then $\mathbf{0}_m$ is in the **range** of T.

28. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** then the **range** and **codomain** are always equal.

Challenge Exercises (Problems)

1. Assume that $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear tranformation and that $T\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, T\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 4 \\ 25 \end{pmatrix}$. Determine $T(\boldsymbol{e}_1), T(\boldsymbol{e}_2)$ and the standard matrix of T.

2. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation** . Give an alternative proof to part 2) of **Theorem** (3.1.4) using the **scalar property** of a linear transformation.

3. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a function. Assume that $T\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, T\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, T\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$. Prove that T cannot be a **linear transformation**.

4. Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^l$ be **linear transformations**. Recall that the **composition** of T and S, denoted by $T \circ S$ is the function from $\mathbb{R}^n$ to $\mathbb{R}^l$ given by $(T \circ S)(\boldsymbol{x}) = T(S(\boldsymbol{x}))$. Prove that $T \circ S$ is a **linear transformation**.

5. Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be **linear transformations**. Define $S + T : \mathbb{R}^n \to \mathbb{R}^m$ by $(S+T)(\boldsymbol{v}) = S(\boldsymbol{v}) + T(\boldsymbol{v})$. Prove that $S+T$ is a **linear transformation**.

6. Do the **induction** for **Theorem** (3.1.2).

Quiz Solutions

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

2. $\begin{pmatrix} 1 \\ 6 \\ 5 \\ 7 \end{pmatrix}$

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

4. $\begin{pmatrix} 0 \\ 7 \\ -4 \\ 2 \end{pmatrix}$

If you had problems with any of these then review **addition** of **n-vectors** and **scalar multiplication**.

3.2. The Product of a Matrix and a Vector

In section (3.1), motivated by **linear transformations** from $\mathbb{R}^n$ to $\mathbb{R}^m$, we defined the **product** of an $m \times n$ matrix A and an **n-vector**, $\boldsymbol{x}$. In this section we study properties of this product.

Am I Ready for This Material

In order to follow this section and master its ideas and methods you need to be familiar with many previously defined concepts. Those that are used extensively in this section are:

linear system

solution

solution set of a linear system

n-vector

$\mathbb{R}^n$

linear combination

span of a sequence of vectors

subspace of $\mathbb{R}^n$

linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$

standard matrix of a linear tranformation

product of an n-vector, x, and an $m \times n$ matrix, A

You will also need to have mastered several methods. Those used frequently in this section are:

Method (1.2.4): How to use matrices to solve a **linear system**

Gaussian elimination

Method (2.3.1): How to determine if an **n-vector** , $\boldsymbol{x}$, belongs to $Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ where $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is a sequence of **n-vectors**.

Quiz

Let $\boldsymbol{v}_1 = \begin{pmatrix} 2 \\ 3 \\ 2 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ -5 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 3 \\ 3 \\ 1 \\ 2 \end{pmatrix}$ and let A be the matrix with columns the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$:

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

1. Find the **solution set** to the **inhomogeneous linear system** with **augmented matrix** $[A| \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}]$.

2. The number of **solutions** to the **homogeneous linear system** with **coefficient matrix** A is

a) no solutions

b) one solution

c) two solutions

d) infinitely many solutions

e) cannot be determined from the given information

3. Determine if $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$.

4. Explain why $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is not all of $\mathbb{R}^4$.

5. Find the **solution set** to the **inhomogeneous linear system** with **augmented matrix** $[A| \begin{pmatrix} 4 \\ 5 \\ 3 \\ -1 \end{pmatrix}]$.

6. Determine if $\begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix}$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$.

Quiz Solutions

New Concepts

the $n \times n$ **identity matrix,** I_n

the **null space of a matrix**

translate of a subset of $\mathbb{R}^n$

affine subspace of $\mathbb{R}^n$

Theory (Why It Works)

In the previous section we defined the concept of a **linear transformation** T from $\mathbb{R}^n$ to $\mathbb{R}^m$. Recall, this is a function $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ which satisfies:

1) $T(\boldsymbol{x} + \boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$ for every pair of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ from $\mathbb{R}^n$; and

2) $T(c\boldsymbol{x}) = cT(\boldsymbol{x})$ for every vector $\boldsymbol{x} \in \mathbb{R}^n$ and scalar c.

We showed how such a **linear transformation** can be encoded by a $m \times n$ matrix defined by

$$A = (T(\boldsymbol{e}_1^n)\; T(\boldsymbol{e}_2^n)\; \ldots T(\boldsymbol{e}_n^n))$$

where $\boldsymbol{e}_i^n$ is the i^{th} **standard basis vector** of $\mathbb{R}^n$. This matrix is called the **standard matrix** of the transformation.

We then used this to motivate the definition of the **product** of an $m \times n$ matrix A with columns $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ by an **n-vector** , $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$. Specifically, the product is given by

$$A\boldsymbol{x} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n.$$

Example 3.2.1. Let

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

Then

$$A\boldsymbol{x} = (-1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} + (1)\begin{pmatrix} 2 \\ 4 \end{pmatrix} + (2)\begin{pmatrix} -1 \\ 1 \end{pmatrix} + (1)\begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Remark 3.2. Alternatively, the i^{th} component of $A\boldsymbol{x}$ can be obtained by taking the i^{th} row of A, which has n entries, multiplying each entry by the corresponding component of $\boldsymbol{x}$, and adding up. In general, if $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$ and $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ then the i^{th} component of $A\boldsymbol{x}$ is

$$x_1 a_{i1} + x_2 a_{i2} + \dots + x_n a_{in}$$

In this section we study this **product** and make use of it in refining our description of the **solution set** of a **linear system**. Before doing so we introduce a particular matrix which plays an important role in the "algebra of matrices," namely the ***identity matrix***.

Definition 3.7. The $n \times n$ ***identity matrix***, I_n, is the matrix with columns $e_1, e_2, \dots, e_n$ the **standard basis vectors** of $\mathbb{R}^n$. Alternatively, it is the matrix with 1's down the main diagonal and 0's every where else. This is the **standard matrix** for the ***identity transformation***, the **linear transformation** $\mathcal{I}_n : \mathbb{R}^n \to \mathbb{R}^n$, which takes every **n-vector** , $\boldsymbol{x}$, to itself: $\mathcal{I}_n(\boldsymbol{x}) = \boldsymbol{x}$.

Example 3.2.2.

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The next theorem highlights some extremely important properties of the matrix product $A\boldsymbol{x}$:

Theorem 3.2.1. *Let A be an $m \times n$ matrix, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ ($n-$vectors) and c a scalar (real number). Then the following hold:*

1. $A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$
2. $A(c\boldsymbol{x}) = c(A\boldsymbol{x})$.
3. $I_n\boldsymbol{x} = \boldsymbol{x}$.

Recall that the transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ given by $T_A(\boldsymbol{x}) = A\boldsymbol{x}$ is a **linear transformation**. In fact, this is how we defined the product $A\boldsymbol{x}$. However, for many readers this may seem like a "proof by magic" and so we give alternative proofs.

1. We demonstrate the proof for a 3×3 matrix, A, and $\boldsymbol{x}, \boldsymbol{y}$ vectors from $\mathbb{R}^3$. Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ denote the columns of A and let

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

Then

$$\boldsymbol{x} + \boldsymbol{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix}.$$

$$A(\boldsymbol{x} + \boldsymbol{y}) = (x_1 + y_1)\boldsymbol{a}_1 + (x_2 + y_2)\boldsymbol{a}_2 + (x_3 + y_3)\boldsymbol{a}_3 =$$

$$x_1\boldsymbol{a}_1 + y_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + y_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3 + y_3\boldsymbol{a}_3 =$$

$$(x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3) + (y_1\boldsymbol{a}_1 + y_2\boldsymbol{a}_2 + y_3\boldsymbol{a}_3) = A\boldsymbol{x} + A\boldsymbol{y}.$$

The remaining parts of the theorem are left as a **challenge exercise**. □

Remark 3.3. We could have begun with a definition of multiplication of an $m \times n$ matrix A by an $n-$vector $\boldsymbol{x}$ and then with the proof of **Theorem** (3.2.1) conclude that the transformation T_A defined by $T_A(\boldsymbol{x}) = A\boldsymbol{x}$ is a **linear transformation**. However, this would have left the definition of **product** of an $m \times n$ matrix and an **n-vector** completely unmotivated and appearing mysterious.

The **matrix product** provides a very powerful tool for further refining how we represent a **linear system**.

Example 3.2.3. Consider the following **linear system**:

$$\begin{array}{rrrcl} x & -2y & -z & = & 4 \\ x & -y & -z & = & 3 \end{array}$$

Instead of two equations we can represent this as a single equation involving two vectors:

$$\begin{pmatrix} x - 2y - z \\ x - y - z \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

On the other hand,

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

$$\begin{pmatrix} 1 & -2 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

and so the **linear system** can be represented by the matrix equation

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

In general, any **linear system** can be written in the form $A\boldsymbol{x} = \boldsymbol{b}$ where A is the **coefficient**
matrix of the system, $\boldsymbol{b}$ is the vector of constant terms and $\boldsymbol{x}$ is the vector of unknowns. This implies the following:

Theorem 3.2.2. *Let A be an $m \times n$ matrix with columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and $\boldsymbol{b}$ an* **m-vector**. *Then the following are equivalent:*

1. *The* **linear system** *with* **augmented matrix** $[A|\boldsymbol{b}]$ *is* **consistent**.
2. *There exists an* **n-vector** $\boldsymbol{x}$ *such that $A\boldsymbol{x} = \boldsymbol{b}$.*
3. *$\boldsymbol{b}$ is a* **linear combination** *of the sequence of columns of A, $(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)$.*
4. *$\boldsymbol{b}$ is in $Span(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)$.*

In order to describe the **solution set** to a **linear system** in a concise fashion it is useful to introduce the following concept:

Definition 3.8. Let A be an $m \times n$ matrix. The ***null space*** of A, denoted $null(A)$, consists of all the **n-vectors** $\boldsymbol{x}$ such that $A\boldsymbol{x} = \boldsymbol{0}$.

It follows from this that $null(A)$ is nothing more than the **solution set** to the
homogeneous linear system with **coefficient matrix** A.

Example 3.2.4. Find $null(A)$ where A is the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 3 & 5 & 4 \\ -2 & 4 & 1 \end{pmatrix}$.

In the usual fashion, we use **Gaussian elimination** to obtain the **reduced echelon form** . We then identify the **leading variables** and **free variables**, assign parameters to the free variables, and express all the variables in terms of the parameters.

$$\begin{pmatrix} 1 & 1 & 1 \\ 3 & 5 & 4 \\ -2 & 4 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 6 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

We write down the **homogeneous linear system** corresponding to this matrix:

$$\begin{array}{ccccc} x & & +\frac{1}{2}z & = & 0 \\ & y & +\frac{1}{2}z & = & 0 \end{array}$$

We have two **leading variables** (x, y) and one **free variable** (z). Setting $z = t$ we get the following paramterization of the **general solution**:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}t \\ -\frac{1}{2}t \end{pmatrix} = t \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}$$

Thus, the **null space of A** is then **span** of the **3-vector** , $\begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}$.

Example 3.2.5. Find the null space of the matrix $C = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 6 \end{pmatrix}$.

To find $null(C)$ we use **Gaussian elimination** to obtain the **reduced echelon form** of C:

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

At this point we can stop. Since every column has a **leading entry**, every variable of the **homogeneous linear system** with C as its **coefficient matrix** is a **leading variable**, there are no **free variables**, and there is only the **trivial solution**. Thus, $null(C) = \{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\}$.

As the last example suggests, for an $m \times n$ matrix A, the **null space of A** consists of only the **zero vector** , $\mathbf{0}_n$, precisely when every column of A is a **pivot column**. In turn, every column of A is a pivot column if and only if the sequence of columns of A is **linearly independent**. These equivalences are so important that we collect them in a theorem:

Theorem 3.2.3. *Let A be an $m \times n$ matrix. Then the following are equivalent:*

1. *The* **null space** *of A consists of just the* **zero vector** *, $null(A) = \{\mathbf{0}_n\}$;*
2. *Every column of A is a* **pivot column***;*
3. *The columns of A are* **linearly independent***.*

In our next theorem we prove that the **null space** of an $m \times n$ matrix A is a **subspace** of $\mathbb{R}^n$.

Theorem 3.2.4. *Let A be an $m \times n$ matrix and assume that $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2$ are in the* **null space** *of A and that c is a scalar. Then*

1. *$\boldsymbol{x}_1 + \boldsymbol{x}_2$ is in $null(A)$.*
2. *$c\boldsymbol{x}$ is in $null(A)$.*

Proof. 1. By assumption $A\boldsymbol{x}_1 = A\boldsymbol{x}_2 = \mathbf{0}_m$. Then

$$A(\boldsymbol{x}_1 + \boldsymbol{x}_2) = A\boldsymbol{x}_1 + A\boldsymbol{x}_2 = \mathbf{0}_m + \mathbf{0}_m = \mathbf{0}_m.$$

The first equality holds by the **additive property of matrix multiplication**. Since $A(\boldsymbol{x}_1 + \boldsymbol{x}_2) = \mathbf{0}$, $\boldsymbol{x}_1 + \boldsymbol{x}_2$ is in $null(A)$.

2. This is left as a **challenge exercise**. $\square$

We have seen for a matrix A that $null(A)$ can be expressed as a **span** of a **linearly independent** sequence of vectors (so the typical vector in $null(A)$ is uniquely a **linear combination** of this sequence). We now deal with the general case of a **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$, equivalently the solutions to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$.

Example 3.2.6. Consider the **linear system**

$$\begin{array}{rrrcr} x & +\ y & +\ z & = & 0 \\ 3x & +\ 5y & +\ 4z & = & -1 \\ -2x & +\ 4y & +\ z & = & -3 \end{array} \tag{3.6}$$

This can be represented as vector equation:

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

In turn this can be represented by the matrix equation:

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

Of course, this can also be represented by the augmented matrix

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

To find the **solution set** to the **linear system** (3.6) we use **Gaussian elimination** to obtain the **reduced echelon form** of the augmented matrix (stopping if at any point we find the augmented column is a **pivot column**).

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

$$\left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{array}\right). \tag{3.7}$$

The matrix in (3.7) is the **augmented matrix** of the **linear system**

$$\begin{array}{ccccccc} x & & & + & \frac{1}{2}z & = & \frac{1}{2} \\ & & y & + & \frac{1}{2}z & = & -\frac{1}{2} \end{array}$$

There are two **leading variables** and one **free variable**, z. We set $z = t$, a parameter, and obtain the solution

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{2} - \frac{1}{2}t \\ -\frac{1}{2} - \frac{1}{2}t \\ t \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{pmatrix} + t \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}.$$

The form of this solution suggests the following definition and notation:

Definition 3.9. Let S be a set of vectors from $\mathbb{R}^n$, for example, the **span** of a sequence of **n-vectors**, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$. Let $\boldsymbol{p}$ be a vector in $\mathbb{R}^n$.

By $\boldsymbol{p} + S = S + \boldsymbol{p}$ we shall mean the set of vectors obtained by adding $\boldsymbol{p}$ to every vector in $S : \boldsymbol{p} + S = \{\boldsymbol{p} + \boldsymbol{x} : \boldsymbol{x} \in S\}$.

We refer to $\boldsymbol{p} + S$ as the ***translate*** of S by $\boldsymbol{p}$. When S is a **subspace** of $\mathbb{R}^n$ then a translate $\boldsymbol{p} + S$ is called an ***affine subspace*** of $\mathbb{R}^n$.

The next theorem demonstrates how we can represent the **solution set** to a **linear system** as an **affine subspace** of $\mathbb{R}^n$ where the **subspace** that is translated is the **null space** of the **coefficient matrix** of the system.

Theorem 3.2.5. *Let A be an $m \times n$ matrix and $\boldsymbol{b}$ an* **m-vector***. Assume that $\boldsymbol{p}$ is a* **solution** *to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$, equivalently, the* **linear system** *with* **augmented matrix** *$[A|\boldsymbol{b}]$. Then the following hold:*

1. If $\boldsymbol{x} = \boldsymbol{p} + \boldsymbol{v}$ with $\boldsymbol{v}$ in $null(A)$ then $A\boldsymbol{x} = \boldsymbol{b}$.

2. If $A\boldsymbol{x} = \boldsymbol{b}$ then there is a unique $\boldsymbol{v}$ in $null(A)$ such that $\boldsymbol{x} = \boldsymbol{p} + \boldsymbol{v}$.

In other words, all the **solutions** are obtained by adding the particular solution $\boldsymbol{p}$ to an element of $null(A)$. Thus, the **solution set** of the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ is $\boldsymbol{p} + null(A)$.

Proof. 1. We first remark that for any **n-vectors** , $\boldsymbol{x}$ and $\boldsymbol{y}$, that $A(\boldsymbol{x}+\boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$ by part 1) of **Theorem** (3.2.1). Since $A\boldsymbol{p} = \boldsymbol{b}$, $A\boldsymbol{v} = \boldsymbol{0}_m$ we have

$$A(\boldsymbol{p}+\boldsymbol{v}) = A\boldsymbol{p} + A\boldsymbol{v} = \boldsymbol{b} + \boldsymbol{0}_m = \boldsymbol{b}.$$

So, indeed, $\boldsymbol{p}+\boldsymbol{v}$ is a solution to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ and the **linear system** with augmented matrix $[A|\boldsymbol{b}]$.

2. Now assume that $A\boldsymbol{x} = \boldsymbol{b}$. We need to find $\boldsymbol{v} \in null(A)$ such that $\boldsymbol{x} = \boldsymbol{p} + \boldsymbol{v}$. Of course, there is precisely one candidate for $\boldsymbol{v}$ namely $\boldsymbol{x} - \boldsymbol{p}$. Clearly, $\boldsymbol{p} + \boldsymbol{v} = \boldsymbol{p} + (\boldsymbol{x} - \boldsymbol{p}) = \boldsymbol{x}$ and is the only vector which satisfies this. We have to show that $\boldsymbol{v}$ is in the **null space** of A. So, we need to multiply $\boldsymbol{v}$ by A and see if we get the **zero vector** , $\boldsymbol{0}_m$. In doing so we can use part 1) of **Theorem** (3.2.1). Thus,

$$A(\boldsymbol{x}-\boldsymbol{p}) = A(\boldsymbol{x}+(-1)\boldsymbol{p}) = A\boldsymbol{x} + A[(-1)\boldsymbol{p}] = A\boldsymbol{x} + (-1)A\boldsymbol{p} = \boldsymbol{b} + (-1)\boldsymbol{b} = \boldsymbol{0}_m$$

as required. $\square$

Remark 3.4. Suppose A is an $m \times n$ matrix, $\boldsymbol{b}$ is an **m-vector** and we are explicitly given the **solution set** to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ as $\boldsymbol{p} + S$ where S is some subset of $\mathbb{R}^n$. Then S must, in fact, be equal to the **null space** of A (this is left as a **challenge exercise**). If S is expressed as the **span** of a **basis** then, in particular, you know the dimension of $null(A)$. This is, in fact, equal to the number of **free variables** in the **homogeneous linear system** with coefficient matrix A, and, in turn, equal to the number of non-pivot columns. Consequently, knowing that A has n columns and the number of non-pivot columns you can deduce the number of **pivot columns** of A.

We conclude with two theorems which collect some equivalent statements for later reference:

Theorem 3.2.6. *Let A be an $m \times n$ matrix. Then the following are equivalent statements:*

1) Every row of A contains a **pivot position**.

2) For every **m-vector** *, $\boldsymbol{b}$, there exists a* **solution** *to the* **linear system** *with* **augmented matrix** $[A|\boldsymbol{b}]$.

3) The sequence of columns of A **spans** $\mathbb{R}^m$.

4) The **column space** *of $A, col(A)$ is equal to $\mathbb{R}^m$.*

5) For every **m-vector** *$\boldsymbol{b}$, the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution.*

Theorem 3.2.7. *Let A be an $m \times n$ matrix. Then the following are equivalent statements:*

1) Every column of A is a **pivot column**.

2) For any **m-vector** *$\boldsymbol{b}$, the* **linear system** *with* **augmented matrix** *$[A|\boldsymbol{b}]$ has at most one* **solution**.

3) The **homogeneous linear system** *with* **coefficient matrix** *A has only the* **trivial solution**.

4) The sequence of columns of A is **linearly independent**.

5) The **null space** *of A is the* **zero subspace** *of $\mathbb{R}^n$, $null(A) = \{\mathbf{0}_n\}$.*

6) The matrix equation $A\boldsymbol{x} = \mathbf{0}_m$ has the unique solution $\boldsymbol{x} = \mathbf{0}_n$.

7) For any **m-vector** *$\boldsymbol{b}$, the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ has at most one solution.*

What You Can Now Do

1. Express a linear system

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

as a matrix equation $A\boldsymbol{x} = \boldsymbol{b}$.

2. Given an $m \times n$ matrix A determine the **null space of A**.

3. Given a **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$ express the **solution set** as a **translate** $\boldsymbol{p} + null(A)$ where $\boldsymbol{p}$ is a particular solution to the system.

Method (How To Do It)

Method 3.2.1. Express a **linear system**

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

as a matrix equation $A\boldsymbol{x} = \boldsymbol{b}$.

Let A be the **coefficient matrix** of the system. Set $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$ then the **linear system** is equivalent to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$.

Example 3.2.7. Express the following **linear system** as a matrix equation:

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

The required matrix equation is

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

Method 3.2.2. Given an $m \times n$ matrix A, find the **null space** of A.

If the matrix is $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$

then finding the **null space** of A is the same as solving the **homogeneous linear system**

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

and this has previously been described. See **Method** (1.2.4).

Example 3.2.8. Find the **null space** of the matrix $A = \begin{pmatrix} 1 & 2 & -4 & -9 \\ 2 & 1 & 1 & 3 \\ 3 & 1 & 3 & 8 \\ -1 & 1 & -5 & -12 \end{pmatrix}$.

Using **Gaussian elimination** we obtain the **reduced echelon form** of A :

$$\to \begin{pmatrix} 1 & 2 & -4 & -9 \\ 0 & -3 & 9 & 21 \\ 0 & -5 & 15 & 35 \\ 0 & 3 & -9 & -21 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -4 & -9 \\ 0 & 1 & -3 & -7 \\ 0 & 1 & -3 & -7 \\ 0 & 1 & -3 & -7 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 2 & 5 \\ 0 & 1 & -3 & -7 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We write out the corresponding homogeneous linear system for the last matrix

$$\begin{array}{ccccccc} x_1 & & + & 2x_3 & + & 5x_4 & = & 0 \\ & x_2 & - & 3x_3 & - & 7x_4 & = & 0 \end{array}$$

There are two **leading variables** (x_1, x_2) and two **free variables** (x_3, x_4). We set the free variables equal to parameters, $x_3 = s, x_4 = t$ and solve for the leading variables in terms of these parameters:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2s-5t \\ 3s+7t \\ s \\ t \end{pmatrix} = \begin{pmatrix} -2s \\ 3s \\ s \\ 0 \end{pmatrix} + \begin{pmatrix} -t \\ t \\ 0 \\ t \end{pmatrix} = s\begin{pmatrix} -2 \\ 3 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -5 \\ 7 \\ 0 \\ 1 \end{pmatrix}$$

Consequently, $null(A)$ is equal to **the span** of the sequence $\left(\begin{pmatrix} -2 \\ 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 7 \\ 0 \\ 1 \end{pmatrix}\right)$:

$$null(A) = Span\left(\begin{pmatrix} -2 \\ 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 7 \\ 0 \\ 1 \end{pmatrix}\right)$$

Method 3.2.3. Given a **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$ express the **solution set** as a **translate** $\boldsymbol{p} + null(A)$ where $\boldsymbol{p}$ is a particular solution to the system.

We use **Method** (1.2.4) to obtain the solution and then represent it in the desired form. Thus, we apply **Gaussian elimination** to obtain the **reduced echelon form** of the augmented matrix $[A|\boldsymbol{b}]$. We write out the corresponding **linear system**, identify the **leading variables** and **free variables**. We set the **free variables** equal to parameters and express the general solution vector in terms of these parameters. This can be written as a sum of a particular solution vector and vectors (one for each free variable) times parameters. Taking the vectors which are multiplied by the parameters we obtain a basis for the **null space** of A.

Example 3.2.9. Find the **solution set** to the matrix equation

$$\begin{pmatrix} 1 & 1 & 1 & -2 \\ 2 & 3 & 1 & -7 \\ -2 & 1 & -5 & -5 \\ 3 & 4 & 2 & -9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -11 \\ 0 \end{pmatrix}$$

We form the augmented matrix $[A|\boldsymbol{b}]$ and apply **Gaussian elimination** to obtain the **reduced echelon form**

$$\begin{pmatrix} 1 & 1 & 1 & -2 & | & 1 \\ 2 & 3 & 1 & -7 & | & -1 \\ -2 & 1 & -5 & -5 & | & -11 \\ 3 & 4 & 2 & -9 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & -2 & | & 1 \\ 0 & 1 & -1 & -3 & | & -3 \\ 0 & 3 & -3 & -9 & | & -9 \\ 0 & 1 & -1 & -3 & | & -3 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 1 & 1 & -2 & | & 1 \\ 0 & 1 & -1 & -3 & | & -3 \\ 0 & 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 2 & 1 & | & 4 \\ 0 & 1 & -1 & -3 & | & -3 \\ 0 & 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$$

We write out the corresponding **linear system**:

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

There are two **leading variables** (x_1, x_2) and two **free variables** $(x_{3,4})$. We set the free variables equal to parameters, $x_3 = s, x_4 = t$ and solve for the leading variables in terms of constants and the parameters:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 - 2s - t \\ -3 + s + 3t \\ s \\ t \end{pmatrix} =$$

$$\begin{pmatrix} 4 \\ -3 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 3 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore the **solution set** is equal to

$$\begin{pmatrix} 4 \\ -3 \\ 0 \\ 0 \end{pmatrix} + Span \left(\begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ 0 \\ 1 \end{pmatrix} \right).$$

Exercises

In exercises 1 and 2 express the **linear system** as a matrix equation. See **Example** (3.2.3) and **Method** (3.2.1).

1.

$$\begin{array}{ccccccc} 2x & - & 3y & - & 4z & = & -1 \\ 3x & + & 2y & + & z & = & 0 \\ & & 3y & - & 2z & = & 4 \\ x & - & y & - & z & = & -3 \end{array}$$

2.

$$\begin{array}{rcrcrcrcr} x_1 & - & x_2 & + & 2x_3 & - & x_4 & = & 0 \\ & & 3x_2 & - & 2x_3 & + & x_4 & = & -1 \\ -x_1 & & & + & 3x_3 & - & x_4 & = & 2 \end{array}$$

In exercises 3 - 8 find the **null space** of the given matrix A. Express $null(A)$ as a **span** of **linearly independent** sequence of vectors. See **Method** (3.2.2).

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

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

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

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

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

8. $A = \begin{pmatrix} 2 & 1 & -1 & 4 & 3 \\ 1 & 2 & 1 & 1 & 2 \\ 0 & 1 & 1 & -1 & 0 \\ -1 & 2 & 3 & -5 & -2 \end{pmatrix}$

In exercises 9 - 16 find the **solution set** to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ for the given matrix A and the constant vector $\boldsymbol{b}$ (which is equivalent to finding the solution set to the **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$. Express your solution in the form $\boldsymbol{p} + S$ for a **subspace** $S = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ and an appropriate **n-vector** $\boldsymbol{p}$. See **Method** (3.2.3).

9. $A = \begin{pmatrix} 1 & 3 & -4 \\ 0 & -1 & 1 \\ 4 & -1 & -3 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 5 \\ -1 \\ 7 \end{pmatrix}$

10. $A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & -1 & 4 \\ 3 & 1 & 1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$

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

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

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

14. $A = \begin{pmatrix} 1 & 3 & -4 \\ 2 & 1 & -3 \\ -2 & 1 & 1 \\ 5 & -3 & -2 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 5 \\ 0 \\ 4 \\ -11 \end{pmatrix}$

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

16. $A = \begin{pmatrix} 1 & 2 & -1 & -2 & 2 \\ 2 & 3 & -4 & -1 & 3 \\ 1 & 3 & 1 & -5 & 3 \\ 1 & 4 & 2 & 0 & 4 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 0 \\ -1 \\ 1 \\ -2 \end{pmatrix}$

In exercises17 - 23 answer true or false and give an explanation.

17. If A is an $m \times n$ **matrix** and $\boldsymbol{x}, \boldsymbol{y}$ are **n-vectors** which satisfy $A\boldsymbol{x} = A\boldsymbol{y}$ then $\boldsymbol{x} = \boldsymbol{y}$.

18. If A is an $m \times n$ **matrix** then $\mathbf{0}_m$ is in $null(A)$.

19. If A is an $m \times n$ **matrix** then $\mathbf{0}_n$ is in $null(A)$.

20. If A is a 2×3 **matrix** and $A\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $A\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ then $A\begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$.

21. If A is 2×3 **matrix** and $A\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $A\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ then $A\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

22. If A is a 5×3 **matrix** then $null(A) \neq \{\mathbf{0}_3\}$.

23. If A is a 3×5 **matrix** then $null(A) \neq \{\mathbf{0}_3\}$.

Challenge Exercises (Problems)

1. Assume that A is a 3×5 matrix and that $null(A)$ is equal to the **span** of $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right)$.

a) Determine the number of **pivot columns** of A. See **Remark** (3.4).

b) If A is in **reduced echelon form** determine A.

2. Assume that A is a 3×4 matrix and $null(A)$ is **spanned by** $\left(\begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 8 \\ -1 \\ 3 \\ 2 \end{pmatrix}\right)$.

a) Determine the number **pivot column** of A. See **Remark** (3.4).

b) If A is in **reduced echelon form** determine A.

3. Let A be a 4×5 matrix. Given the following information about $null(A)$ determine if the columns of A are a **spanning sequence** for $\mathbb{R}^4$. See **Remark** (3.4).

a) $null(A) = Span(\begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \\ -3 \end{pmatrix})$

b) $null(A) = Span(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix})$

c) $null(A) = \{\mathbf{0}_5\}$.

d) $null(A) = Span(\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 3 \\ -4 \\ 5 \end{pmatrix})$.

4. Assume that the matrix A is 4×6. Explain why it is not possible that $null(A)$ is equal to the **span** of the 6-vector $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$.

5. Assume the **solution set** to the matrix equation $A\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ is $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) + \boldsymbol{p}$ where $\boldsymbol{v}_1 = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix}$, $\boldsymbol{v}_2 = \begin{pmatrix} -1 \\ 3 \\ 0 \\ 1 \end{pmatrix}$ and $\boldsymbol{p} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}$.

a) What is the size of A?

b) Compute $A[3\begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix} - 2\begin{pmatrix} -1 \\ 3 \\ 0 \\ 1 \end{pmatrix} + 4\begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}]$. See **Theorem** (3.2.1).

c) If A is in **reduced echelon form** determine A.

6. Explain why it is not possible that the **solution set** to the matrix equation $A\boldsymbol{x} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}$ is $Span(\boldsymbol{v}) + \boldsymbol{p}$ where $\boldsymbol{v} = \begin{pmatrix} 2 \\ -1 \\ 1 \\ -3 \\ 1 \end{pmatrix}$ and $\boldsymbol{p} = \begin{pmatrix} 1 \\ -1 \\ 3 \\ -1 \\ 3 \end{pmatrix}$.

7. Complete the proof of **Theorem** (3.2.1).

8. Complete the proof of **Theorem** (3.2.4).

Quiz Solutions

1. The system is **inconsistent** so there are no **solutions** and the **solution set** of the **linear system** is empty. The augmented matrix $\left(\begin{array}{cccc|c} 2 & 3 & 0 & 3 & 0 \\ 3 & 4 & 1 & 3 & 0 \\ 2 & 2 & 2 & 1 & 0 \\ -1 & 1 & -5 & 2 & 1 \end{array}\right)$ has **reduced echelon form** $\left(\begin{array}{cccc|c} 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right)$. Not right, see **Method** (1.2.4).

2. d) infinitely many solutions since not every column of A is a **pivot column** as can be deduced from the **reduced echelon form** of A. Not right, see **Theorem** (1.2.4).

3. $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ is not in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ since by 1) the **inhomogeneous linear system** with **augmented matrix** $[A| \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}]$ is **inconsistent**. Not right, see **Theorem** (2.2.2).

4. Since $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ is not in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ it follows that $Span(\boldsymbol{v}_2, \dots, \boldsymbol{v}_4) \neq \mathbb{R}^4$.

5. The **solution set** is

$$\{ \begin{pmatrix} 2-3t \\ -1+2t \\ t \\ 1 \end{pmatrix} | t \in \mathbb{R} \}.$$

Not right, see **Method** (2.3.1).

6. Yes. $\begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix}$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ since the system $[A| \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix}]$ is **consistent**. The **solution set** is equal to $\{ \begin{pmatrix} -3t \\ 1+2t \\ t \\ -1 \end{pmatrix} | t \in \mathbb{R} \}$. In particular, $\begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \boldsymbol{v}_2 - \boldsymbol{v}_4$.

Not right, see **Method** (2.3.1).

3.3. Matrix Addition and Multiplication

In this section we define a sum, scalar product and product of matrices (when they have compatible sizes) and begin to develop the algebra of matrices - the rules that are satisfied by these operations.

Am I Ready for This Material

In order to follow this section and master its ideas and methods you need to be familiar with many previously defined concepts. Those that are used extensively in this section are:

$\mathbb{R}^n$

matrix

rows and columns of a matrix

a **linear transformation** T from $\mathbb{R}^n$ to $\mathbb{R}^m$

product of an $m \times n$ matrix and an n-vector

$n \times n$ identity matrix, I_n

standard matrix of a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$

Quiz

In each of the following **multiply** the matrix and the vector.

1. $\begin{pmatrix} 1 & 2 & 4 \\ -2 & 4 & 16 \\ 4 & 6 & 10 \\ 0 & 3 & 9 \end{pmatrix} \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}$

2. $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \\ -4 \end{pmatrix}$

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

4. $\begin{pmatrix} 2 & -3 & 0 & 1 \\ 4 & 2 & 1 & -2 \\ -1 & 1 & 3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

5. $\begin{pmatrix} 1 & -1 & 2 & 3 \\ 2 & 4 & -1 & 5 \\ -3 & 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 3 \\ -3 \end{pmatrix}$

Quiz Solutions

New Concepts

The concepts introduced in this section are:

row matrix of length n

column matrix of length m

$m \times n$ **zero matrix**

square matrix

main diagonal of a square matrix

diagonal matrix

upper and lower triangular matrix

equality of matrices

matrix addition

scalar product of a matrix

negative of a matrix

product of two matrices

transpose of a matrix

symmetric matrix

Theory (Why It Works)

Since we are elaborating on one of the fundamental objects of linear algebra we begin with several definitions and then give examples.

Definition 3.10. Recall that a ***matrix*** is a rectangular array of numbers. If the number of rows is m and the number of columns is n then we refer to it as an $m \times n$ matrix. The general form of such a matrix is

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

The i^{th} ***row*** of the matrix $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$ is $\begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix}$.

The j^{th} ***column*** of this matrix is $\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$.

The $(i,j)^{th}$ ***entry*** is a_{ij}.

Example 3.3.1. The matrix $\begin{pmatrix} 1 & -1 & 0 & 2 & 4 \\ -2 & 3 & -6 & 0 & -5 \\ 7 & -9 & 11 & -4 & 8 \end{pmatrix}$ is 3×5. The (2,4)-entry is 0, the (3,5)-entry is 8.

Definition 3.11. A $1 \times n$ matrix is a ***row matrix*** of length n. An $m \times 1$ matrix is a ***column matrix*** of length m.

We can view a matrix as being composed of its columns and at times it will be useful to do so. In this way an $m \times n$ matrix corresponds to a sequence of n vectors from **m-space**, $\mathbb{R}^m$.

Definition 3.12. Any matrix all of whose entries are zero is called a ***zero matrix***. If it is an $m \times n$ matrix we will denote it by $\mathbf{0}_{m \times n}$.

Example 3.3.2. The 3×2 zero matrix is $\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$.

The 3×4 zero matrix, $\mathbf{0}_{3\times 4} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

Definition 3.13. An $n \times n$ matrix is said to be a ***square matrix***. We will denote the collection of all $n \times n$ matrices by $M_{n\times n}(\mathbb{R})$.

Suppose A is an $n \times n$ matrix. The entries $a_{ii}, i = 1, 2, \ldots, n$ are called ***diagonal entries*** and the sequence $(a_{11}, a_{22}, \ldots, a_{nn})$ is the the ***main diagonal***.

If the only non-zero entries of an $n \times n$ matrix A lie on the main diagonal then A is called a ***diagonal matrix***. If A is an $n \times n$ diagonal matrix with main diagonal $(a_1, \ldots, a_n)$ we will use the notation $A = diag(a_1, \ldots, a_n)$.

Example 3.3.3. The **main diagonal** of $\begin{pmatrix} -2 & 3 & -1 & 0 \\ 3 & 3 & -7 & -5 \\ 6 & -\frac{1}{2} & \frac{5}{6} & -1 \\ 0 & 0 & -4 & \frac{2}{5} \end{pmatrix}$ is $(-2, 3, \frac{5}{6}, \frac{2}{5})$.

The matrix $A = \begin{pmatrix} -3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -9 \end{pmatrix}$ is a 3×3 **diagonal matrix**. $A = diag(-3, 5, -9)$.

Definition 3.14. An ***upper triangular matrix*** is a **square matrix** with the property that all the entries below the **main diagonal** are zero. A ***lower triangular matrix*** is a square matrix with the property that all the entries above the main diagonal are zero.

Example 3.3.4. The matrix $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$ is upper triangular. On the other hand, the matrix $\begin{pmatrix} 0 & 0 & 0 & 0 \\ -3 & 2 & 0 & 0 \\ -2 & 1 & -1 & 0 \\ -1 & 0 & -2 & -1 \end{pmatrix}$ is lower triangular. The matrix $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ is neither upper nor lower triangular.

Definition 3.15. Two matrices are ***equal*** when they have the same size and corresponding entries are equal.

Operations on Matrices

So far all we have been doing is naming things. We now introduce some further notions and begin to develop the "**algebra**" of matrices.

Definition 3.16. When A and B are two matrices of the same size $(m \times n)$ then we define the ***sum***, $A + B$, to be the $m \times n$ matrix whose $(i, j)^{th}$ **entry** is the sum of the corresponding entries of A and B for each for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

In symbols we write: if $A = (a_{ij})$ and $B = (b_{ij})$ then $A + B = (a_{ij} + b_{ij})$.

Remark 3.5. This definition is, hopefully, the obvious one. It can be motivated by **linear transformations**. Let the S and T be linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ have **standard matrices** A and B. Define $S + T : \mathbb{R}^n \to \mathbb{R}^m$ by $(S + T)(\boldsymbol{x}) = S(\boldsymbol{x}) + T(\boldsymbol{x})$. Then $S + T$ is a **linear transformation** (see **Challenge Exercise 2**). Since it is a **linear transformation** it has a **standard matrix**, C. By the definition of the standard matrix, the j^{th} column of C is $(S + T)(\boldsymbol{e}_j) = S(\boldsymbol{e}_j) + T(\boldsymbol{e}_j)$. However, $S(\boldsymbol{e}_i)$ is the i^{th} column of A and $T(\boldsymbol{e}_i)$ is the i^{th} column of B. This implies that $C = A + B$.

Example 3.3.5.

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

$$\begin{pmatrix} -2 & 2 & 0 & 5 \\ -2 & 3 & 2 & 4 \end{pmatrix}.$$

Definition 3.17. Let A be an $m \times n$ matrix and c a scalar. The ***scalar product*** of c and A, denoted by cA, is the $m \times n$ matrix whose **entries** are obained by multiplying the entries of A by the scalar c.

In symbols we write: If $A = (a_{ij})$ then $cA = (ca_{ij})$.

Of special importance for each matrix A is the particular scalar multiple $(-1)A$ which we denote by $-A$. This is the ***negative*** of A.

Remark 3.6. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** and c is a scalar, let $cT : \mathbb{R}^n \to \mathbb{R}^m$ to be the transformation defined by

$$(cT)(\boldsymbol{v}) = cT(\boldsymbol{v}).$$

Then cT is a **linear transformation** (see **Challenge Exercise 3**). Let A be the **standard matrix** of T. Since cT is a linear transformation it has a standard matrix, which turns out to be cA. To see this, let $\boldsymbol{e}_j^n$ be the j^{th} **standard basis vector** of $\mathbb{R}^n$, a_j the j^{th} column of A and let a'_j be the j^{th} column of the standard matrix of cT. Then by the definition of the standard matrix, $a_j = T(\boldsymbol{e}_j^n)$ and $a'_j = (cT)(\boldsymbol{e}_j^n)$. However, $(cT)(\boldsymbol{e}_j^n) = c[T(\boldsymbol{e}_j^n)] = ca_j$.

Properties of matrix addition and scalar multiplication

The next result is omnibus with many parts. Fundamentally, it asserts that the collection of $m \times n$ matrices with the operations of **matrix addition** and **scalar multiplication of matrices** is a vector space, a concept we introduce in Chapter 5.

Theorem 3.3.1. *Let A, B, C be $m \times n$ matrices and a, b, c be any scalars. Then the following hold:*

*1. $(A + B) + C = A + (B + C)$. **Matrix addition** is **associative**.*
*2. $A + B = B + A$. **Matrix addition** is **commutative**.*
*3. $A + \mathbf{0}_{m\times n} = A$. The $m \times n$ **zero matrix** is an **identity element** **matrix addition***
*4. $A + (-A) = 0_{m\times n}$. The **negative of a matrix** is an **additive inverse** relative to the **zero matrix**.*
*5. $c(A + B) = cA + cB$. **Scalar multiplication of matrices** **distributes** over sums.*
*6. $(a + b)C = aC + bC$. Another form of the **distributive property**.*
7. $(ab)C = a(bC)$.
8. $1A = A$.
9. $0A = 0 - m \times n$.

Proof. The proofs are straightforward but all depend on representing the matrices with arbitrary elements. For example, look at the first one. On the left hand side we have to first compute $A + B$. If the **(i, j) entry** of A is a_{ij} and the (i, j) entry of B is b_{ij} then the (i, j) entry of $A + B$ is $a_{ij} + b_{ij}$. We must then add $A + B$ to C. The (i, j) entry of C is c_{ij}. Therefore the (i, j) entry of $(A + B) + C$ is $(a_{ij} + b_{ij}) + c_{ij}$. The exact same reasoning yields that the (i, j) entry of $A + (B + C)$ is $a_{ij} + (b_{ij} + c_{ij})$. However, these are equal since the associative law of addition holds in the reals.

The second one is true since it depends entirely on the commutative law of addition holding in every entry but this a property of addition of real numbers. The rest of the arguments are similar.

□

Matrix Multiplication

We introduce one further operation, that of ***matrix multiplication***. This was anticipated in the previous section where we defined the **product** of an $m \times n$ **matrix** A and an **n-vector** , $\boldsymbol{x}$. We now define the concept of matrix multiplication in full generality.

Definition 3.18. Let B be a $l \times m$ matrix and A be an $m \times n$ matrix. Let the sequence of columns of A be $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ Then BA is defined to be the matrix whose j^{th} column is $B\boldsymbol{a}_j$, $BA = (B\boldsymbol{a}_1 \ \ldots \ B\boldsymbol{a}_n)$.

Remark 3.7. By this definition BA has n columns which are each $l \times 1$ and consequently the matrix BA is a $l \times n$ matrix. Thus, the product of a $l \times m$ matrix and an $m \times n$ matrix is a $k \times n$ matrix.

There is another way in which one can view the product, called the ***row-column computation***. Of course, you get the same result. This method computes the individual entries of BA when B is $l \times m$ and A is $m \times n$.

To find the $(i, j)-$**entry** of BA multiply the entries of the i^{th} row of B (which is an $1 \times m$ matrix) by the corresponding entries of the j^{th} column of A (which is an $m \times 1$ matrix) and add these up.

Example 3.3.6. Let's compute BA where $B = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 0 & 1 \end{pmatrix}$ and $A = \begin{pmatrix} -3 & 1 \\ 2 & 4 \\ -1 & 0 \end{pmatrix}$.

According to this procedure the **(1,1)-entry** is the result of multiplying the entries of the first row of B by the corresponding entries of the first column of A and adding up. Thus, we get

(1)(-3) + (-2)(2) + (3)(-1) = (-3) + (-4) + (-3) = -10

Likewise we can compute the other entries:

(1,2): (1)(1) + (-2)(4) + (3)(0) = -7

(2,1): (2)(-3) + (0)(2) + (1)(-1) = -7.

(2,2): (2)(1) + (0)(4) + (1)(0) = 2 + 0 + 0 = 2.

Thus, $BA = \begin{pmatrix} -10 & -7 \\ -7 & 2 \end{pmatrix}$

Remark 3.8. The definition of **matrix multiplication** may appear strange; however, it can be motivated in the following way: Let $S : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation** with standard matrix A and $T : \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations with standard matrix B. Further, let the columns of A be $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$. Then by the definition of the standard matrix $\boldsymbol{a}_j = S(\boldsymbol{e}_j^n), \boldsymbol{e}_j$ the j^{th} **standard basis vector** of $\mathbb{R}^n$.

The **composition** of S and $T, T \circ S : \mathbb{R}^n \to \mathbb{R}^k$ is defined by $(T \circ S)(\boldsymbol{x}) = T(S(\boldsymbol{x}))$. This is a **linear transformation** (See **Challenge Exercise 4**). Consequently, it has a **standard matrix**, C. By the definition of the standard matrix, the j^{th} column of C is $(T \circ S)(\boldsymbol{e}_j^n)$.

By the definition of the composition $T \circ S$ we have $(T \circ S)(\boldsymbol{e}_j^n) = T(S(\boldsymbol{e}_j^n)) = T(\boldsymbol{a}_j)$. However, the **standard matrix** B for T satisfies $T(\boldsymbol{x}) = B\boldsymbol{x}$ for every $\boldsymbol{x} \in \mathbb{R}^m$. In particular, $T(\boldsymbol{a}_j) = B\boldsymbol{a}_j$ and therefore the j^{th} column of C is $B\boldsymbol{a}_j$, that is, $C = BA$ by **Definition** (3.18). Thus, the **standard matrix** of a composition of T and S is the product of the standard matrix of T and the standard matrix of S. This is the motivation for this peculiar definition.

In a number of important ways which we will shortly discuss and illustrate, **matrix multiplication** is very different from ordinary multiplication of numbers and therefore you cannot apply all your intuitions. However in many ways matrix multiplication and multiplication of numbers share basic properties. We collect and summarize these properties in a comprehensive theorem:

Theorem 3.3.2. *Let A be an $m \times n$ matrix. Then the following hold:*

1. Let B be an $n \times p$ matrix and C a $p \times q$ matrix. Then $(AB)C = A(BC)$. This is the **associative law of multiplication for matrices**.

2. Let B and C be $n \times p$ matrices. Then $A(B + C) = AB + AC$. This is the **left distributive law for matrix multiplication**.

3. Let B and C be $k \times m$ matrices. Then $(B + C)A = BA + CA$. This is the **right distributive law for matrix multiplication**.

4. Let B be an $l \times m$ matrix and c a scalar. Then $c(BA) = (cB)A = B(cA)$.

5. $I_m A = A I_n = A$. This says the **identity matrix** *is a* **multiplicative identity**.

Proof. We prove the first two and leave the remainder as **Challenge Exercises**.

1. Suppose first that $C = \boldsymbol{v}$ is a $p \times 1$ **column matrix**, with components $v_1, v_2, \dots v_p$. Then B is an $n \times p$ matrix. Assume that $B = (\boldsymbol{b}_1 \boldsymbol{b}_2 \dots \boldsymbol{b}_p)$.

For the purpose of exposition let's take $p = 3$. Therefore, $B = (\boldsymbol{b}_1\boldsymbol{b}_2\boldsymbol{b}_3)$, $\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$,

and $AB = (A\boldsymbol{b}_1\ A\boldsymbol{b}_2\ A\boldsymbol{b}_3)$. Then $(AB)\boldsymbol{v} = \begin{pmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & A\boldsymbol{b}_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} =$

$v_1(A\boldsymbol{b}_1) + v_2(A\boldsymbol{b}_2) + v_3(A\boldsymbol{b}_3) = A(v_1\boldsymbol{b}_1) + A(v_2\boldsymbol{b}_2) + A(v_3\boldsymbol{b}_3) =$

$A(v_1\boldsymbol{b}_1 + v_2\boldsymbol{b}_2 + v_3\boldsymbol{b}_3) = A(B\boldsymbol{v})$ and so the result holds in this case.

The general case for p holds in exactly the same way. So assume we have proved that $(AB)\boldsymbol{v} = A(B\boldsymbol{v})$ when B is an $n \times p$ matrix and $\boldsymbol{v}$ is a $p \times 1$ matrix. Now assume that C is a $p \times q$ matrix and view C as composed of its columns. Again, for the purposes of exposition let's suppose C has 4 columns $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3, \boldsymbol{c}_4$ so that $C = (\boldsymbol{c}_1\ \boldsymbol{c}_2\ \boldsymbol{c}_3\ \boldsymbol{c}_4)$

Now $(AB)C = (AB)(\boldsymbol{c}_1\ \boldsymbol{c}_2\ \boldsymbol{c}_3\ \boldsymbol{c}_4) = ((AB)\boldsymbol{c}_1\ (AB)\boldsymbol{c}_2\ (AB)\boldsymbol{c}_3\ (AB)\boldsymbol{c}_4) =$

$(A(B\boldsymbol{c}_1)\ A(B\boldsymbol{c}_2)\ A(B\boldsymbol{c}_3)\ A(B\boldsymbol{c}_4)) = A(B\boldsymbol{c}_1\ B\boldsymbol{c}_2\ B\boldsymbol{c}_3\ B\boldsymbol{c}_4) = A(BC)$.

2. For purposes of exposition let us assume that B and C are $n \times 3$:

$B = (\boldsymbol{b}_1\ \boldsymbol{b}_2\ \boldsymbol{b}_3), C = (\boldsymbol{c}_1\ \boldsymbol{c}_2\ \boldsymbol{c}_3)$

Now $AB = (A\boldsymbol{b}_1\ A\boldsymbol{b}_2\ A\boldsymbol{c}_3), AC = (A\boldsymbol{c}_1\ A\boldsymbol{c}_2\ A\boldsymbol{c}_3)$.

Then $AB + AC = ([A\boldsymbol{b}_1 + A\boldsymbol{c}_1]\ [A\boldsymbol{b}_2 + A\boldsymbol{c}_2]\ [A\boldsymbol{b}_3 + A\boldsymbol{c}_3])$

On the other hand, $B + C = (\boldsymbol{b}_1 + \boldsymbol{c}_1\ \boldsymbol{b}_2 + \boldsymbol{c}_2\ \boldsymbol{b}_3 + \boldsymbol{c}_3)$ and then

$A(B + C) = (A[\boldsymbol{b}_1 + \boldsymbol{c}_1]\ A[\boldsymbol{b}_2 + \boldsymbol{c}_2]\ A[\boldsymbol{b}_3 + \boldsymbol{c}_3])$

Now each $\boldsymbol{b}_i, \boldsymbol{c}_i$ is a column and we previously demonstrated that for an $m \times n$ matrix A and $n \times 1$ column matrices $\boldsymbol{x}, \boldsymbol{y}$ that $A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$.

This applies in the present situation and therefore for each $i = 1, 2, 3$ we have $A(\boldsymbol{b}_i + \boldsymbol{c}_i) = A\boldsymbol{b}_i + A\boldsymbol{c}_i$ and this completes the proof.

$\square$

Remark 3.9. One could use properties of **linear transformation** and **standard matrices of linear transformations** to prove most of these. For example, it is not difficult to prove that the composition of functions is always associative (this is **Challenge Exercise 5**) and the **associativity** of **matrix multiplication** can be proved from this. Similarly, for the **distributive properties**. However, this would probably make the proofs look like magic or simply tricks.

Example 3.3.7. For $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 0 & -1 \\ 3 & 4 \end{pmatrix}, C = \begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix}$,

verify each of the following:

1) $(AB)C = A(BC)$;

2) $A(B+C) = AB + AC$.

1) $(AB)C = [\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 3 & 4 \end{pmatrix}]\begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 7 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} -5 & 45 \\ -2 & 24 \end{pmatrix}$.

$A(BC) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}[\begin{pmatrix} 0 & -1 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix}] = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & -3 \\ -2 & 24 \end{pmatrix} = \begin{pmatrix} -5 & 45 \\ -2 & 24 \end{pmatrix}$.

2) $AB + AC = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 7 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 0 & 10 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 17 \\ 4 & 7 \end{pmatrix}$.

$A(B+C) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}(\begin{pmatrix} 0 & -1 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} -2 & 4 \\ 1 & 3 \end{pmatrix}) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -2 & 3 \\ 4 & 7 \end{pmatrix} = \begin{pmatrix} 6 & 17 \\ 4 & 7 \end{pmatrix}$.

How matrix multiplication differs from real multiplication

I. Commutativity does not hold

The most obvious way in which **matrix multiplication** differs from real number multiplication is that it is not **commutative**, that is, it is not always the case that for two matrices A and B that $AB = BA$. First of all, this may fail since it is not always the case that if we can **multiply** AB that we can also multiply BA. Hence there may be no way to compare AB with BA least of all that they are **equal**. For example, if A is 2×3 and B is 3×4.

On the other hand, it may be possible to compute the **products** AB and BA but these matrices may have different **sizes** and therefore cannot be **equal**. As an example, if A is 2×3 and B is 3×2. Then AB is 2×2 and BA is 3×3.

Finally, it can be the case that A and B are **square**, say 3×3. Then both AB and BA are square of the same size, but it still may be that $AB \neq BA$. We illustrate with an example.

Example 3.3.8.

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

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

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

II. Cancellation does not hold for matrix multiplication

It can also be the case that two matrics are non-zero but their **product** is the **zero matrix**. For example:

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

$$AB = \begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix} \begin{pmatrix} 1 & -3 \\ 2 & -6 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Related to this "peculiarity" of **matrix multiplication** is the fact that it is possible to have matrices A, B, C such that $AB = AC$ but $B \neq C$. The example of above shows that $AB = A\mathbf{0}_{22}$ but $B \neq \mathbf{0}_{22}$.

For another example, take

$$A = \begin{pmatrix} 2 & -3 \\ -4 & 6 \end{pmatrix}, B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, C = \begin{pmatrix} 4 & -2 \\ 3 & -1 \end{pmatrix}$$

Then $AB = AC$ but $B \neq C$.

How the collection of square matrices is like the real numbers

If we take the collection of **square matrices of size** $n, M_{n\times n}$, then we can **add** any two and get an $n \times n$ matrix and all the usual rules pertaining to addition hold:

1. **Associativity**

2. **Communativity**

3. Existence of an **identity element,** $\mathbf{0}_{n\times n}$

4. Existence of an **additive inverse**: for each $n \times n$ matrix A the element $-A$ satisfies $A + (-A) = 0$.

Also, we can **multiply** any two $n \times n$ matrices and get back an $n \times n$ matrix. Moreover, this multiplication satisifes:

1. **Associativity**

2. **Existence of an identity element,** I_n.

3. The **right and left distributive laws** hold.

The transpose of a matrix

We introduce one final concept that will be an important tool used in the next section when we study invertibility.

Definition 3.19. For an $m \times n$ matrix A the ***transpose***, which is denoted by A^{Tr}, is defined to be the matrix obtained from A by interchanging its **rows and columns**. Therefore, A^{Tr} is an $n \times m$ matrix. The $(i,j)-$**entry** of A^{Tr} is the $(j,i)-$entry of A.

Example 3.3.9. a) $\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}^{Tr} = \begin{pmatrix} 1 & -2 & 3 \end{pmatrix}$.

b) $\begin{pmatrix} 3 & -1 & 0 & 2 \\ 1 & 2 & 3 & -4 \end{pmatrix}^{Tr} = \begin{pmatrix} 3 & 1 \\ -1 & 2 \\ 0 & 3 \\ 2 & -4 \end{pmatrix}$.

c) $\begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}^{Tr} = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$.

d) $\begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -5 \\ 3 & -5 & 6 \end{pmatrix}^{Tr} = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -5 \\ 3 & -5 & 6 \end{pmatrix}$.

e) $\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 3 \\ -2 & -3 & 0 \end{pmatrix}^{Tr} = \begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -3 \\ 2 & 3 & 0 \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 3 \\ -2 & -3 & 0 \end{pmatrix}$.

Definition 3.20. A **square matrix** A such that $A^{Tr} = A$ is called a ***symmetric matrix***. A square matrix A which satisfies $A^{Tr} = -A$ is a ***skew symmetric matrix***.

Thus, the matrix A in part d) of **Example** (3.3.9) is **symmetric** while the matrix of part e) is **skew symmetric**.

We prove some properties of the transpose.

Theorem 3.3.3. *1.* $(A^{Tr})^{Tr} = A$.

2. $(A + B)^{Tr} = A^{Tr} + B^{Tr}$.

3. For a scalar c, $(cA)^{Tr} = cA^{Tr}$.

4. $(AB)^{Tr} = B^{Tr}A^{Tr}$.

Proof. The first three are parts are quite straightforward and so we concentrate on the last part.

4. For the purposes of exposition let's assume that A is 2×3 and B is 3×3. Say,

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

Then

$$B^{Tr} = \begin{pmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \\ b_{13} & b_{23} & b_{33} \end{pmatrix}, A^{Tr} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}.$$

Now the **(j, i)-entry** of $(AB)^{Tr}$ is the $(i, j)-$entry of AB and this is just the number obtained when the i^{th} **row** of A is multiplied by the j^{th} **column** of B. For example, the (3,1)-entry of $(AB)^{Tr}$ is the (1,3)-entry of AB and this is the product of the first row of A with the third column of B :

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \end{pmatrix} \begin{pmatrix} b_{13} \\ b_{23} \\ b_{33} \end{pmatrix} = a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33}.$$

On the other hand, the (3,1)-entry of $B^{Tr}A^{Tr}$ is the product of the third **row** of B^{Tr} with the first **column** of A^{Tr}. But the third row of B^{Tr} is the third column of B and the first column of A^{Tr} is the first row of A : It therefore follows that

$$\begin{pmatrix} b_{13} & b_{23} & b_{33} \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = b_{13}a_{11} + b_{23}a_{12} + b_{33}a_{13}$$

so, indeed, the **(3,1)-entry** of $(AB)^{Tr}$ is equal to the (3,1)-entry of $B^{Tr}A^{Tr}$. This holds for every entry of the matrices $(AB)^{Tr}$ and $B^{Tr}A^{Tr}$. $\square$

What You Can Now Do

1. Given two $m \times n$ **matrices**, A and B, compute their **sum**.

2. Given an $m \times n$ **matrix** A, and a scalar c, compute the **scalar product**, cA.

3. Given an $m \times n$ **matrix** A and an $l \times m$ matrix B and natural numbers $i \leq m, j \leq n$ compute the $(i,j)-$**entry** of the matrix BA.

4. Given an $m \times n$ **matrix** A and an $l \times m$ matrix B compute the **matrix product** BA.

Method (How To Do It)

Method 3.3.1. Given two $m \times n$ **matrices**, A and B, compute their **sum**.

If the **matrices** are $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \dots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}$ then the sum and of A and B is given by adding the corresponding **entries** of the matrices, $A + B =$

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

Example 3.3.10. Let $A = \begin{pmatrix} 1 & -2 & 0 & 4 \\ 2 & 0 & 5 & 3 \\ -5 & 7 & 4 & -4 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 4 & -5 & 0 \\ -6 & 0 & 1 & -3 \\ -1 & 1 & -2 & 3 \end{pmatrix}$.

Then $A + B = \begin{pmatrix} 4 & 2 & -5 & 4 \\ -4 & 0 & 6 & 0 \\ -6 & 8 & 2 & -1 \end{pmatrix}$ and $A - B = \begin{pmatrix} -2 & -6 & 5 & 4 \\ 8 & 0 & 4 & 6 \\ -4 & 6 & 6 & -7 \end{pmatrix}$.

Method 3.3.2. Given an $m \times n$ **matrix** A, and a scalar c, compute the **scalar product**, cA.

Multiply each of the entries of A by c. So, if $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$ then

$$cA = \begin{pmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \dots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{pmatrix}.$$

Example 3.3.11. Let $A = \begin{pmatrix} 1 & -2 & 0 & 4 \\ 2 & 0 & 5 & 3 \\ -5 & 7 & 4 & -4 \end{pmatrix}$. Compute $-A, 2A, \frac{1}{2}A, -3A$.

$$-A = \begin{pmatrix} -1 & 2 & 0 & -4 \\ -2 & 0 & -5 & -3 \\ 5 & -7 & -4 & 4 \end{pmatrix}, 2A = \begin{pmatrix} 2 & -4 & 0 & 8 \\ 4 & 0 & 10 & 6 \\ -10 & 14 & 8 & -8 \end{pmatrix}$$

$$\frac{1}{2}A = \begin{pmatrix} \frac{1}{2} & -1 & 0 & 2 \\ 1 & 0 & \frac{5}{2} & \frac{3}{2} \\ -\frac{5}{2} & \frac{7}{2} & 2 & -2 \end{pmatrix}, -3A = \begin{pmatrix} -3 & 6 & 0 & -12 \\ -6 & 0 & -15 & -9 \\ 15 & -21 & -12 & 12 \end{pmatrix}.$$

Method 3.3.3. Given an $m \times n$ **matrix** A and an $l \times m$ matrix B and natural numbers $i \leq m, j \leq n$ compute the $(i, j)-$**entry** of the matrix BA.

The $(i, j)-$entry of BA is computed by multiplying the entries of the i^{th} **row** of B by the corresponding entries of the j^{th} **column** of A, and adding the resulting projects. The number obtained is the $(i, j)-$entry of BA.

Example 3.3.12. Let $B = \begin{pmatrix} 1 & -1 & 2 & 0 & -3 \\ 3 & -2 & 5 & -4 & 6 \\ 2 & -1 & 4 & 3 & -1 \\ 4 & 5 & -3 & -2 & 0 \end{pmatrix}, A = \begin{pmatrix} 2 & -4 & 6 \\ 1 & 3 & 5 \\ 0 & -2 & 3 \\ -1 & 4 & -8 \\ 3 & 2 & 1 \end{pmatrix}.$

Since B is 4×5 and A is 5×3 it is possible to multiply B and A and the resulting matrix is 4×3. Compute the (1,1), (2,3) and (4,2) entries.

The (1,1) entry is $(1 \ -1 \ 2 \ 0 \ -3) \begin{pmatrix} 2 \\ 1 \\ 0 \\ -1 \\ 3 \end{pmatrix} =$

$1 \times 2 + (-1) \times 1 + 2 \times 0 + 0 \times (-1) + (-3) \times 3 = 2 - 1 - 9 = -8.$

The (2,3) entry is $(3 \ -2 \ 5 \ -4 \ 6) \begin{pmatrix} 6 \\ 5 \\ 3 \\ -8 \\ 1 \end{pmatrix} =$

$3 \times 6 + (-2) \times 5 + 5 \times 3 + (-4) \times (-8) + 6 \times 1 = 18 - 10 + 15 + 32 + 6 = 61.$

the (4,2) entry is $(4\ 5\ \ -3\ \ -2\ 0)\begin{pmatrix} -4 \\ 3 \\ -2 \\ 4 \\ 2 \end{pmatrix} =$

$4 \times (-4) + 5 \times 3 + (-3) \times (-2) + (-2) \times 4 + 0 \times 2 = -16 + 15 + 6 - 8 = -3.$

Method 3.3.4. Given an $m \times n$ **matrix** A and an $l \times m$ matrix B compute the **matrix product** BA.

The **matrix product** BA by definition is $(B\boldsymbol{a}_1 B\boldsymbol{a}_2 \dots B\boldsymbol{a}_n)$ where the sequence of columns of A is $(\boldsymbol{a}_1, \boldsymbol{a}_2, \dots \boldsymbol{a}_n)$.

Example 3.3.13. Let $B = \begin{pmatrix} 1 & -1 & 2 & 0 & -3 \\ 3 & -2 & 5 & -4 & 6 \\ 2 & -1 & 4 & 3 & -1 \\ 4 & 5 & -3 & -2 & 0 \end{pmatrix}$ and $A = \begin{pmatrix} 2 & -4 & 6 \\ 1 & 3 & 5 \\ 0 & -2 & 3 \\ -1 & 4 & -8 \\ 3 & 2 & 7 \end{pmatrix}$.

Compute BA.

$$B\boldsymbol{a}_1 = \begin{pmatrix} 1 & -1 & 2 & 0 & -3 \\ 3 & -2 & 5 & -4 & 6 \\ 2 & -1 & 4 & 3 & -1 \\ 4 & 5 & -3 & -2 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \\ -1 \\ 3 \end{pmatrix} =$$

$$2\begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix} + 1\begin{pmatrix} -1 \\ -2 \\ -1 \\ 5 \end{pmatrix} + 0\begin{pmatrix} 2 \\ 5 \\ 4 \\ -3 \end{pmatrix} + (-1)\begin{pmatrix} 0 \\ -4 \\ 3 \\ -2 \end{pmatrix} + 3\begin{pmatrix} -3 \\ 6 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -8 \\ 26 \\ -3 \\ 15 \end{pmatrix}.$$

$$B\boldsymbol{a}_2 = \begin{pmatrix} 1 & -1 & 2 & 0 & -3 \\ 3 & -2 & 5 & -4 & 6 \\ 2 & -1 & 4 & 3 & -1 \\ 4 & 5 & -3 & -2 & 0 \end{pmatrix} \begin{pmatrix} -4 \\ 3 \\ -2 \\ 4 \\ 2 \end{pmatrix} =$$

$$(-4)\begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix} + 3\begin{pmatrix} -1 \\ -2 \\ -1 \\ 5 \end{pmatrix} + (-2)\begin{pmatrix} 2 \\ 5 \\ 4 \\ -3 \end{pmatrix} + 4\begin{pmatrix} 0 \\ -4 \\ 3 \\ -2 \end{pmatrix} + 2\begin{pmatrix} -3 \\ 6 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -17 \\ -32 \\ -9 \\ -3 \end{pmatrix}.$$

$$B\boldsymbol{a}_3 = \begin{pmatrix} 1 & -1 & 2 & 0 & -3 \\ 3 & -2 & 5 & -4 & 6 \\ 2 & -1 & 4 & 3 & -1 \\ 4 & 5 & -3 & -2 & 0 \end{pmatrix} \begin{pmatrix} 6 \\ 5 \\ 3 \\ -8 \\ 7 \end{pmatrix} = 6 \begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix} + 5 \begin{pmatrix} -1 \\ -2 \\ -1 \\ 5 \end{pmatrix} + 3 \begin{pmatrix} 2 \\ 5 \\ 4 \\ -3 \end{pmatrix} +$$

$$(-8) \begin{pmatrix} 0 \\ -4 \\ 3 \\ -2 \end{pmatrix} + 7 \begin{pmatrix} -3 \\ 6 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -14 \\ 97 \\ -12 \\ 56 \end{pmatrix}.$$

Thus $Ba = \begin{pmatrix} -8 & -17 & -14 \\ 26 & -32 & 97 \\ -3 & -9 & -12 \\ 15 & -3 & 56 \end{pmatrix}$.

Example 3.3.14. Let $B = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 3 & 4 \\ 3 & 4 & 2 \end{pmatrix}$ and $A = \begin{pmatrix} -10 & 16 & -7 \\ 8 & -13 & 6 \\ -1 & 2 & -1 \end{pmatrix}$. Then

$$B \begin{pmatrix} -10 \\ 8 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 3 & 4 \\ 3 & 4 & 2 \end{pmatrix} \begin{pmatrix} -10 \\ 8 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

$$B \begin{pmatrix} 16 \\ -13 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 3 & 4 \\ 3 & 4 & 2 \end{pmatrix} \begin{pmatrix} 16 \\ -13 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

$$B \begin{pmatrix} -7 \\ 6 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 3 & 4 \\ 3 & 4 & 2 \end{pmatrix} \begin{pmatrix} -7 \\ 6 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore $BA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. This example will be better understood after the next section.

Exercises

In exercises 1-5 compute the **sum** of the given matrices. See **Method** (3.3.1).

1. $\begin{pmatrix} 2 & -3 \\ -4 & 7 \end{pmatrix}, \begin{pmatrix} -3 & 5 \\ 2 & -1 \end{pmatrix}$

2. $\begin{pmatrix} 3 & 0 \\ -5 & 8 \\ 2 & 6 \end{pmatrix}, \begin{pmatrix} -3 & 0 \\ 5 & -8 \\ -2 & -6 \end{pmatrix}.$

3. $\begin{pmatrix} 3 & -2 & 5 \\ -2 & 7 & 4 \\ 4 & 5 & 11 \end{pmatrix}, \begin{pmatrix} -2 & 3 & -1 \\ 3 & -5 & 1 \\ -1 & 1 & -6 \end{pmatrix}.$

4. $\begin{pmatrix} 1 & 2 & 4 & 6 \\ -2 & 3 & 5 & -7 \\ 6 & -3 & -5 & 0 \end{pmatrix}, \begin{pmatrix} -3 & -5 & 2 & -2 \\ 0 & 4 & -6 & 7 \\ -5 & 8 & 3 & -2 \end{pmatrix}.$

5. $\begin{pmatrix} 1 & -3 \\ 2 & 6 \\ -4 & 9 \\ 5 & 11 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$

In exercises 6-11, each pair consisting of the given matrix A and scalar c compute the **scalar product, cA**. See **Method** (3.3.2).

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

7. $\begin{pmatrix} 4 & 5 \\ -6 & -2 \\ -1 & 5 \end{pmatrix}, c = -1$

8. $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, c = 3$

9. $\begin{pmatrix} 2 & 1 & -3 \\ 3 & -2 & 1 \\ -4 & 7 & 9 \end{pmatrix}, c = 0$

10. $\begin{pmatrix} 0 & 0 & 0 \\ 3 & -1 & 1 \end{pmatrix}, c = 10$

11. $\begin{pmatrix} 3 & -1 \\ -4 & 1 \end{pmatrix}, c = -2$

In exercises 12-14 compute the (1,1) and (2,1) **entries** of the **product** of the given matrices. See **Method** (3.3.3).

12. $\begin{pmatrix} -3 & 1 \\ 2 & 4 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & -2 & 3 \\ 2 & 0 & 1 \end{pmatrix}$

13. $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 & 2 & -1 \\ 0 & 3 & -1 & 4 \end{pmatrix}$

14. $\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}, \begin{pmatrix} -2 & 1 \\ 2 & 5 \\ -1 & -3 \end{pmatrix}$

In 15-20 compute the **product** of the two given matrices. See **Method** (3.3.4).

15. $\begin{pmatrix} 1 & -2 & 2 & 4 \\ 0 & 3 & 5 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 & -3 \\ 1 & -1 & 2 \\ -3 & 5 & -1 \\ 3 & 1 & 2 \end{pmatrix}.$

16. $\begin{pmatrix} 1 & 2 & 3 & 1 \\ 2 & 5 & 7 & 4 \\ 1 & 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 1 & 7 \\ 1 & -2 & -4 & -3 \\ -1 & 0 & 2 & -1 \\ 0 & 1 & 1 & 2 \end{pmatrix}$

17. $\begin{pmatrix} 1 & 3 & 2 \\ 2 & 5 & 5 \\ 1 & 5 & 1 \end{pmatrix} \begin{pmatrix} 20 & -7 & -5 \\ -3 & 1 & 1 \\ -5 & 2 & 1 \end{pmatrix}.$

18. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 5 \\ 2 & -4 \\ 5 & 7 \end{pmatrix}$

19. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ -5 & -4 & -3 & -2 \\ 0 & 1 & 2 & 4 \end{pmatrix}$

20. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & -2 \\ 3 & 0 & -2 \\ 3 & 7 & -5 \end{pmatrix}.$

In exercises 21-23 for the given matrices A, B compute $A^{Tr}, B^{Tr}, BA, (BA)^{Tr}$ and $A^{Tr}B^{Tr}$.

21. $B = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}, A = \begin{pmatrix} 0 & 4 \\ -4 & 0 \end{pmatrix}$

22. $B = \begin{pmatrix} 1 & 2 \\ -2 & 3 \\ -3 & 5 \end{pmatrix}, A = \begin{pmatrix} 2 & 4 & -5 \\ -4 & 1 & 0 \end{pmatrix}$.

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

In exercises 24 - 28 answer true or false and give an explanation.

24. If A and B are 2×2 **matrices** and $A \neq B$ then the **matrix products** AB and BA are not equal.

25. If A is a 2×2 **matrix** and $A^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ then $A = \mathbf{0}_{2\times 2}$.

26. A **diagonal matrix** is always **symmetric**.

27. If A, B and C are $m \times n$ **matrices** and $A + B = A + C$ then $B = C$.

28. If A and B are 2×2 **matrices** and $BA = \mathbf{0}_{2\times 2}$ then $AB = \mathbf{0}_{2\times 2}$.

Challenge Exercises (Problems)

1. Consider a matrix

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

Explain what is wrong with the following argument that $a = b = 0$.

If $a = b = 0$ then we are done. Suppose not both $a = 0$ and $b = 0$. Then the matrix

$$\begin{pmatrix} b & -a \\ b & -a \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

However, the **matrix product**

$$\begin{pmatrix} b & -a \\ b & -a \end{pmatrix} \begin{pmatrix} a & a \\ b & b \end{pmatrix} = \begin{pmatrix} ba + (-a)b & ba + (-a)b \\ ba + (-a)b & ba + (-a)b \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

But then we must have

$$\begin{pmatrix} a & a \\ b & b \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

hence $a = b = 0$ after all.

2. Let $S, T : \mathbb{R}^n \to \mathbb{R}^m$ be two **linear transformations**. Define $S + T : \mathbb{R}^n \to \mathbb{R}^m$ by $(S+T)(\boldsymbol{x}) = S(\boldsymbol{x}) + T(\boldsymbol{x})$. Prove that $S + T$ is a linear transformation.

3. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation** and c a scalar. Define $cT : \mathbb{R}^n \to \mathbb{R}^m$ by

$$(cT)(\boldsymbol{v}) = cT(\boldsymbol{v})$$

for all vectors $\boldsymbol{v} \in \mathbb{R}^n$. Prove that cT is a **linear transformation**.

4. Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^l$ be **linear transformations**. Prove that the **composition**, $T \circ S : \mathbb{R}^n \to \mathbb{R}^l$ is a linear transformation.

5. Let $f : Y \to Z, g : X \to Y$ and $h : W \to X$ be **functions**. Prove that **composition** is associative, that is, $f \circ (g \circ h) = (f \circ g) \circ h$.

6. Let $R, S : \mathbb{R}^n \to \mathbb{R}^m$ be **linear transformations** and $T : \mathbb{R}^m \to \mathbb{R}^k$ be a linear transformation.

a) Prove that $T \circ (R + S) = (T \circ R) + (T \circ S)$.

b) Let A be the **standard matrix** of R, B the standard matrix of S and C the standard matrix of T. Prove that $C(A + B) = CA + CB$.

7. Let A be an $m \times n$ matrix, B be an $l \times m$ matrix, and c a scalar. Prove that $c(BA) = (cB)A = B(cA)$.

8. Let A be an $m \times n$ matrix. Prove that the sequence of columns of A **spans** $\mathbb{R}^m$ if and only if the sequence of columns of A^{Tr} is **linearly independent**.

Quiz Solutions

1. $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$. Not right, see **Method** (3.1.4).

2. $\begin{pmatrix} -3 \\ -9 \\ -29 \end{pmatrix}$. Not right, see **Method** (3.1.4).

3. $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. Not right, see **Method** (3.1.4).

4. $\begin{pmatrix} 0 \\ 5 \\ 8 \end{pmatrix}$. Not right, see **Method** (3.1.4).

5. $\begin{pmatrix} 0 \\ -18 \\ -11 \end{pmatrix}$. Not right, see **Method** (3.1.4).

3.4. Invertible Matrices

This section is about square matrices which have the property that it is possible to express the identity matrix as a product of the given matrix and another matrix, called its inverse.

Am I Ready for This Material

Previous concepts or procedures which will be used in this section are:

linear system

homogeneous linear system

matrix in echelon form

matrix in reduced echelon form

an echelon form of a matrix

the reduced echelon form of a matrix

pivot positions of a matrix

pivot columns of a matrix

n-vector

$\mathbb{R}^n$

subspace of $\mathbb{R}^n$

span of a sequence of n-vectors

linearly independent sequence of n-vectors

product of an $m \times n$ matrix and an n-vector

null space of a matrix

product of an $l \times m$ matrix and an $m \times n$ matrix

the $n \times n$ identity matrix

There are several procedures you will need to have mastered; these are

Gaussian elimination

Method (1.2.4) for solving a linear system

Method (2.2.2) for expressing a vector as a linear combination of a sequence of vectors

Method (2.4.1) for determining if a sequence of vectors is **linearly independent**

Quiz

Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & -4 \\ -3 & -7 & 2 \end{pmatrix}$

1. Verify that the sequence of columns of A **spans** $\mathbb{R}^3$.

2. Verify that the sequence of columns of A is **linearly independent**.

3. Explain why you can conclude that there are vectors $\boldsymbol{v}_i, i = 1, 2, 3$, such that $A\boldsymbol{v}_i = \boldsymbol{e}_i^3$ where $\boldsymbol{e}_i^3, i = 1, 2, 3$, are the **standard basis vectors** for $\mathbb{R}^3$.

4. For each $i = 1, 2, 3$ solve the **linear system** with **augmented matrix** $[A|\boldsymbol{e}_i^3]$ to find the vectors $\boldsymbol{v}_i$.

5. Set $\mathcal{B} = (\boldsymbol{v}_1 \boldsymbol{v}_2 \boldsymbol{v}_3)$. What matrix is AB?

6. Find the product BA.

Quiz Solutions

New Concepts

In this section we introduce some important concepts for **square matrices**, in particular, the notion of an **invertible matrix**. As we shall see the following questions are all equivalent for an $n \times n$ matrix A

Is A **invertible**?

Is the sequence of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ n-vectors **linearly independent**?

Does the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ of n-vectors **span** $\mathbb{R}^n$?

The other concepts introduced are:

inverse of a non-singular matrix

singular or non-invertible matrix

left inverse to an $m \times n$ matrix A

right inverse to an $m \times n$ matrix A

Theory (Why It Works)

Consider a **linear system** of n equations in n variables with **augmented matrix** $[A|\boldsymbol{b}]$ where A is an $n \times n$ matrix and $\boldsymbol{b}$ is a **n-vector**. By **Theorem** (3.2.2) this is equivalent to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$.

Suppose we can find an $n \times n$ matrix B such that $BA = I_n$, where I_n is the $n \times n$ **identity matrix**.

If there is a solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$ then it will also satisfy $B(A\boldsymbol{x}) = B\boldsymbol{b}$. However, $B(A\boldsymbol{x}) = (BA)\boldsymbol{x} = I_n\boldsymbol{x} = \boldsymbol{x}$.

Therefore, if there is a solution, it must be $B\boldsymbol{b}$. Thus, equipped with such a matrix B we can use it to solve the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ for an arbitrary $\boldsymbol{b}$.

Example 3.4.1. Consider the **linear system**

$$\begin{array}{ccccccc} 2x & - & 3y & + & 6z & = & 7 \\ x & - & y & + & 2z & = & 2 \\ 3x & - & 2y & + & 3z & = & 1 \end{array} \tag{3.8}$$

Alternatively, the system can be represented by the **augmented matrix**

$$\begin{pmatrix} 2 & -3 & 6 & | & 7 \\ 1 & -1 & 2 & | & 2 \\ 3 & -2 & 3 & | & 1 \end{pmatrix}$$

and yet another way by the matrix equation

$$\begin{pmatrix} 2 & -3 & 6 \\ 1 & -1 & 2 \\ 3 & -2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix}.$$

Let

$$B = \begin{pmatrix} -1 & 3 & 0 \\ -3 & 12 & -2 \\ -1 & 5 & -1 \end{pmatrix}.$$

We compute BA :

$$BA = \begin{pmatrix} -1 & 3 & 0 \\ -3 & 12 & -2 \\ -1 & 5 & -1 \end{pmatrix} \begin{pmatrix} 2 & -3 & 6 \\ 1 & -1 & 2 \\ 3 & -2 & 3 \end{pmatrix} = I_3.$$

As a consequence, we can compute a **solution** to the system (3.8):

$$\begin{pmatrix} -1 & 3 & 0 \\ -3 & 12 & -2 \\ -1 & 5 & -1 \end{pmatrix} \begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}.$$

Check that this is a **solution**.

Remark 3.10. We claim that the **product** AB is also the **identity matrix**:

$$AB = \begin{pmatrix} 2 & -3 & 6 \\ 1 & -1 & 2 \\ 3 & -2 & 3 \end{pmatrix} \begin{pmatrix} -1 & 3 & 0 \\ -3 & 12 & -2 \\ -1 & 5 & -1 \end{pmatrix} = I_3.$$

So, indeed, AB is also the **identity matrix, I_3**. We shall show later in this section if A is a **square matrix of size n** and there is an $n \times n$ matrix B such that $BA = I_n$ then also $AB = I_n$, though the proof of this is not entirely trivial. In any case, this example motivates the following definition:

Definition 3.21. We shall say that an $n \times n$ matrix A is ***invertible or non-singular*** if there is a matrix B such that $AB = BA = I_n$. The matrix B is called an ***inverse*** to A.

Sometimes we will only be interested in multiplying on one side to get an **identity matrix** and the next definition deals with this situation:

Definition 3.22. Let A be an $m \times n$ matrix. If B is an $n \times m$ matrix and $BA = I_n$ then we say B is a ***left inverse*** of A. On the other hand, if $AB = I_m$ then we say B is a ***right inverse***.

Making use of these definitions, the claim made in **Remark** (3.10) can be restated as follows: if A is a **square matrix** and B is a **left inverse** to A then B is also a **right inverse** and hence an **inverse** to A.

An immediate question which comes to mind is: Can an $n \times n$ **matrix** A have more than one **inverse**? We answer this in the next result.

Theorem 3.4.1. *If the $n \times n$ **matrix** A has an **inverse** then it is unique.*

We will actually prove a stronger result from which **Theorem** (3.4.1) will immediately follow. The result is:

Theorem 3.4.2. *Let A be an $n \times n$ **matrix** and assume there are $n \times n$ matrices B, C such that $BA = AC = I_n$. Then $B = C$.*

Proof. **Matrix multiplication** is associative by **Theorem** (3.3.2). In particular, $B(AC) = (BA)C$. Since we are assuming that $AC = I_n$ it follows that $B(AC) = BI_n = B$. On the other hand, since $BA = I_n$ we can conclude that $(BA)C = I_nC = C$. Thus, $B = C$ as claimed. □

We now prove **Theorem** (3.4.1): an **invertible matrix** has a unique **inverse**. Thus, assume A is an invertible matrix and both B and C are inverses. Then B is a left inverse, C is a right inverse and therefore by **Theorem** (3.4.2) we conclude that $B = C$. □

Remark 3.11. When A is an **invertible matrix** we will denote the unique **inverse** to A by the notation A^{-1}.

Definition 3.23. If the **square matrix** A does not have an **inverse** then we will say that A is ***non-invertible or singular***.

Shortly, we will describe a procedure for determining if a **square matrix** A has an **inverse** or not and how to compute it when it does. However, there is a very simple criterion for determining whether a 2×2 matrix has an inverse and, when it does, for constructing it. This is the subject of the next theorem:

Theorem 3.4.3. *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *Then* A *is* **invertible** *if and only if* $ad - bc \neq 0$. *If* $ad - bc \neq 0$ *then* $A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Proof. Suppose $ad - bc \neq 0$. We check by **multiplying** that $\begin{pmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$ is the **inverse** to $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} = \begin{pmatrix} \frac{ad-bc}{ad-bc} & 0 \\ 0 & \frac{ad-bc}{ad-bc} \end{pmatrix} = I_2.$$

A similar calculation shows that the product in the other order is also the **identity matrix**.

On the other hand, assume that A has an inverse B but $ad - bc = 0$. Clearly not all a, b, c, d are zero since otherwise we would have $I_2 = BA = B\mathbf{0}_{2\times 2} = \mathbf{0}_{2\times 2}$. However, a simple calculation like the one above shows that

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

But then

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = I_2\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = (BA)\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} =$$

$$B(A\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}) = B\mathbf{0}_{2\times 2} = \mathbf{0}_{2\times 2}.$$

Consequently, $a = b = c = d = 0$, a contradiction. $\square$

As suggested by **Example** (3.4.1) the **invertibility of a matrix** A has consequences for finding a **solution** to a **linear system** with **coefficient matrix** A. We make this explicit in the following.

Theorem 3.4.4. *Let* A *be an* **invertible** $n \times n$ **matrix**. *Then for each* **n-vector** , $\boldsymbol{b}$, *there is one and only one* **solution** *to the* **linear system** *which is equivalent to the matrix equation* $A\boldsymbol{x} = \boldsymbol{b}$.

Proof. Suppose that $\boldsymbol{x}$ is a **solution**, that is, $A\boldsymbol{x} = \boldsymbol{b}$. Multiplying both sides on the left by A^{-1} yields $A^{-1}(A\boldsymbol{x}) = A^{-1}\boldsymbol{b}$. However, $A^{-1}(A\boldsymbol{x}) = (A^{-1}A)\boldsymbol{x} = I_n\boldsymbol{x} = \boldsymbol{x}$. Therefore. $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ and so if there is a solution it is unique.

On the other hand, $A(A^{-1}\boldsymbol{b}) = (AA^{-1})\boldsymbol{b} = I_n\boldsymbol{b}$ and therefore, indeed, $A^{-1}\boldsymbol{b}$ is a **solution**. □

There are some important consequences of this result:

Corollary 3.4.5. *Let A be an* **invertible matrix** *with sequence of columns $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$. Then $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is a* **spanning sequence of $\mathbb{R}^n$**.

Proof. There exists a solution to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\boldsymbol{b}$ is in the **span** of the sequence of columns of A. However, by **Theorem** (3.4.4), for every vector $\boldsymbol{b} \in \mathbb{R}^n$ there is an $\boldsymbol{x} \in \mathbb{R}^n$ such that $A\boldsymbol{x} = \boldsymbol{b}$. Consequently, every **n-vector** $\boldsymbol{b}$ is in the **span** of $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ so that $Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \mathbb{R}^n$. □

In our next result we prove that the sequence of columns of an invertible matrix is **linearly independent**.

Corollary 3.4.6. *Assume that A is an* **invertible $n \times n$ matrix**. *Then the sequence of columns of A is* **linearly independent**.

Proof. By **Theorem** (3.4.4) there is only one solution to the matrix equation $A\boldsymbol{x} = \boldsymbol{0}_n$. However, $\boldsymbol{0}_n$ is such a solution. Consequently, $null(A) = \{\boldsymbol{0}_n\}$. By **Theorem** (3.2.3) it follows that the sequence of columns of A is **linearly independent**. □

We next explore some properties of the **inverse** of an **invertible matrix**.

Theorem 3.4.7.

1. If A and B are $n \times n$ **invertible matrices** *then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. Alternatively, the product of* **invertible matrices** *is invertible and the inverse of the product is the product of the inverses, in reverse order.*

2. If A is **invertible** *then so is A^{-1} and $(A^{-1})^{-1} = A$.*

Proof. 1. Simply multiply: $(AB)(B^{-1}A^{-1}) = A[B(B^{-1}A^{-1})] = A[(BB^{-1}A^{-1}] = A[I_nA^{-1}] = AA^{-1} = I_n$.

On the other hand, $(B^{-1}A^{-1})(AB) = B^{-1}[A^{-1}(AB)] = B^{-1}[(A^{-1}A)B] = B^{-1}[I_nB] = B^{-1}B = I_n$.

2. Since $AA^{-1} = A^{-1}A$ it follows that A^{-1} is **invertible** with **inverse** equal to A. □

Remark 3.12. 1) By the **principle of mathematical induction** the first part of **Theorem** (3.4.7) can be extended to any finite number of **invertible** $n \times n$ matrices:

If $A_1, A_2, \dots, A_k$ are **invertible** $n \times n$ matrices then $A_1 A_2 \dots A_k$ is invertible and $(A_1 A_2 \dots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \dots A_2^{-1} A_1^{-1}$.

2) The converse to part 1) of **Theorem** (3.4.7) holds and we prove it at the end of the section.

Notation

When A is invertible we can take powers of A^{-1} and in this way define A^m for any integer m. Thus, we define for a natural number (positive integer) k

$A^{-k} = (A^{-1})^k = A^{-1}A^{-1} \dots A^{-1}$ where there are k factors in the multiplication.

These powers of a matrix satisfy the usual rules for exponents. We state these without a proof:

Theorem 3.4.8. *Let A be an* **invertible** *$n \times n$ matrix. Let k, l be integers.*

1. A^k is invertible and $(A^k)^{-1} = (A^{-1})^k = A^{-k}$.

2. $A^k A^l = A^{k+l}$.

3. $(A^k)^l = A^{kl}$.

When a matrix is **invertible** we can cancel it in multiplications like in ordinary arithmetic:

Theorem 3.4.9. *Assume that the $n \times n$ matrix A is* **invertible**.

1. If $AB = AC$ then $B = C$.

2. If $BA = CA$ then $B = C$.

Proof. We prove the first part; the second is proved in exactly the same way.

Assume that $AB = AC$. Multiply both sides on the left by A^{-1} :

$B = I_n B = (A^{-1}A)B = A^{-1}(AB) = A^{-1}(AC) = (A^{-1}A)C = I_n C = C.$ $\square$

Enough abstraction for now, let's see how to test whether a matrix is **invertible** and how to compute an **inverse** to a matrix when it exists. We begin with an example.

Example 3.4.2. Let $A = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 1 & -1 \\ 3 & 1 & -2 \end{pmatrix}$. Suppose B is an **inverse** to A and the sequence of columns of B is $(\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ so that $B = (\boldsymbol{b}_1\ \boldsymbol{b}_2\ \boldsymbol{b}_3)$.

Then by the **definition of the product**, $AB = (A\boldsymbol{b}_1\ A\boldsymbol{b}_2\ A\boldsymbol{b}_3) = I_3 = \begin{pmatrix} \boldsymbol{e}_1^3 & \boldsymbol{e}_2^3 & \boldsymbol{e}_3^3 \end{pmatrix}$ where $\boldsymbol{e}_1^3, \boldsymbol{e}_2^3, \boldsymbol{e}_3^3$ are the **standard basis vectors** of $\mathbb{R}^3$. So, it appears that to find the **inverse** we have to solve three **linear systems** (equivalently, matrix equations):

$$\begin{aligned} A\boldsymbol{x} &= \boldsymbol{e}_1^3 \\ A\boldsymbol{x} &= \boldsymbol{e}_2^3 \\ A\boldsymbol{x} &= \boldsymbol{e}_3^3. \end{aligned}$$

Of course, we solve these by application of **Gaussian elimination** . However, we do not have to do these as three separate problems. We know that if the inverse exists that in each instance we will get a unique solution. That means when we apply **Gaussian elimination** to any of the augmented matrices $[A|\boldsymbol{e}_i^3], i = 1, 2, 3$, each column of A is a **pivot column**. Since A is 3×3 this must mean that the **reduced echelon form** is I_3. The steps of that we perform in the three applications of **Gaussian elimination** will be the same and, consequently, we can do all three at once and find the **inverse** in one sequence of computations as follows:

Adjoint all three columns, $\boldsymbol{e}_1^3, \boldsymbol{e}_2^3, \boldsymbol{e}_3^3$ to A to get the augmented matrix

$$[A|I_3] = \begin{pmatrix} 3 & 2 & 0 & | & 1 & 0 & 0 \\ 2 & 1 & -1 & | & 0 & 1 & 0 \\ 3 & 1 & -2 & | & 0 & 0 & 1 \end{pmatrix}$$

Use **Gaussian elimination** to obtain the **reduced echelon form** . If A is **invertible** then on the left hand side A will be transformed to the 3×3 **identity matrix,** I_3. The right hand side will be A^{-1}. Let's do it.

The steps of the **Gaussian elimination** are as follows (though I diverge from the usual algorithm by subtracting the second row from the first to get a a 1 in the (1,1) position without introducing fractions)

$$\to \begin{pmatrix} 1 & 1 & 1 & | & 1 & -1 & 0 \\ 2 & 1 & -1 & | & 0 & 1 & 0 \\ 3 & 1 & -2 & | & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & | & 1 & -1 & 0 \\ 0 & -1 & -3 & | & -2 & 3 & 0 \\ 0 & -2 & -5 & | & -3 & 3 & 1 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 0 & -2 & | & -1 & 2 & 0 \\ 0 & 1 & 3 & | & 2 & -3 & 0 \\ 0 & 0 & 1 & | & 1 & -3 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & 1 & -4 & 2 \\ 0 & 1 & 0 & | & -1 & 6 & -3 \\ 0 & 0 & 1 & | & 1 & -3 & 1 \end{pmatrix}.$$

Thus the **inverse** of A is $\begin{pmatrix} 1 & -4 & 2 \\ -1 & 6 & -3 \\ 1 & -3 & 1 \end{pmatrix}$.

We check the calculation:

$$\begin{pmatrix} 3 & 2 & 0 \\ 2 & 1 & -1 \\ 3 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & -4 & 2 \\ -1 & 6 & -3 \\ 1 & -3 & 1 \end{pmatrix} = I_3.$$

Check the other multiplication also results in I_3.

Below we summarize the method for determining whether a square ($n \times n$) matrix has an inverse and how to compute the inverse when it exists:

Matrix inversion algorithm

Assume A is an $n \times n$ matrix. To find A^{-1} if it exists:

1. Form the augmented matrix $[A|I_n]$.

2. Apply **Gaussian elimination** to $[A|I_n]$. If at any point a **zero row** appears on the left hand side, **STOP**, the matrix A is **not invertible**. Otherwise, carry out the **Gaussian elimination** until the matrix is in **reduced echelon form** . The left hand side will be the $n \times n$ **identity matrix, I_n**, and so this form will be $[I_n|B]$. The matrix B appearing on the right hand side of the **reduced echelon form** is A^{-1}.

Remark 3.13. At this point all we actually know is that B is a **right inverse** of $A, AB = I_n$. We do not yet know that $BA = I_n$. We proceed to prove this and then collect some conditions which are equivalent to a matrix being **invertible**.

Theorem 3.4.10. *Let A be an $m \times n$ matrix. Then the following are equivalent:*

*1. The sequence of columns of A **spans** $\mathbb{R}^m$;*
2. There is a matrix B such that $AB = I_m$.

Proof. Assume that the sequence of columns of A **spans** $\mathbb{R}^m$. Then for every vector $\boldsymbol{b} \in \mathbb{R}^m$ there is a solution to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$. In particular this holds if if $\boldsymbol{b} = \boldsymbol{e}_j^m$, the j^{th} **standard basis vector**. Let $\boldsymbol{u}_j$ be a solution, $A\boldsymbol{u}_j = \boldsymbol{e}_j^m$, and set $B = (\boldsymbol{u}_1\ \boldsymbol{u}_2\ \ldots\ \boldsymbol{u}_m)$ then

$$AB = (A\boldsymbol{u}_1\ A\boldsymbol{u}_2\ \ldots\ A\boldsymbol{u}_m) = (\boldsymbol{e}_1^m\ \boldsymbol{e}_2^m\ \ldots\ \boldsymbol{e}_m^m) = I_m.$$

Conversely, assume that there is a matrix B such that $AB = I_m$. Let $\boldsymbol{b}$ be an **m-vector**. Set $\boldsymbol{x} = B\boldsymbol{b}$. Then $A\boldsymbol{x} = A(B\boldsymbol{b}) = (AB)\boldsymbol{b} = I_m\boldsymbol{b} = \boldsymbol{b}$. Therefore, for every $\boldsymbol{b} \in \mathbb{R}^m$ the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution and this is equivalent to the span of the sequence of columns of A is all of $\mathbb{R}^m$. $\square$

Remark 3.14. If A is $m \times n$ and A has a **right inverse** it follows from **Theorem** (2.3.2) that $n \geq m$.

We next prove a similar result for the existence of a **left inverse**:

Theorem 3.4.11. *Let A be an $m \times n$ matrix and assume that A has a* **left inverse** *B. Then* **the null space of A** *is the* **zero subspace***, $null(A) = \{\mathbf{0}_n\}$.*

Proof. Assume $A\boldsymbol{x} = \mathbf{0}_n$. Then $B(A\boldsymbol{x}) = B\mathbf{0}_n = \mathbf{0}_n$. On the other hand $B(A\boldsymbol{x}) = (BA)\boldsymbol{x} = I_n\boldsymbol{x} = \boldsymbol{x}$. Therefore $\boldsymbol{x} = \mathbf{0}_n$. $\square$

Remark 3.15. If A is an $m \times n$ matrix and A has a **left inverse** then by **Theorem** (2.4.2) it follows that $n \leq m$.

We now prove a theorem that validates the **matrix inversion algorithm**:

Theorem 3.4.12. *Let A be an $n \times n$ matrix. Then the following are equivalent:*

1. *A is* **invertible***;*
2. *A has a* **right inverse***.*

Proof. If A is **invertible** then A^{-1} is both a **right and left inverse** and there is nothing to prove.

Conversely, assume that B is a **right inverse** to A. By **Theorem** (3.4.10) the sequence of columns of A **spans** $\mathbb{R}^n$. On the other hand, A is a **left inverse** to B. By (3.4.11) we can conclude that the **the null space of B** is the **zero subspace** of $\mathbb{R}^n, null(B) = \{\mathbf{0}_n\}$, equivalently, the sequence of columns of B is **linearly independent**. However, B is a square matrix. By **Theorem** (2.4.4) this implies that the sequence of columns of B **spans** $\mathbb{R}^n$. By **Theorem** (3.4.10) B has a **right inverse** C. Since $AB = BC = I_n$ it follows by **Theorem** (3.4.2) that $A = C$ so that also $BA = I_n$ and B is an **inverse** to A. $\square$

Before closing this section with a summary of the many conditions that are equivalent to the **invertibility** of an $n \times n$ matrix we consider the relationship between the invertibility of an $n \times n$ matrix and its **transpose**.

Theorem 3.4.13. *Assume an $n \times n$ matrix A is* **invertible***. Then the* **transpose of A** *is invertible and $(A^{Tr})^{-1} = (A^{-1})^{Tr}$. In words, the inverse of the transpose of A is the transpose of the inverse of A.*

Proof. We previously saw that $(AB)^{Tr} = B^{Tr}A^{Tr}$. Since $AA^{-1} = I_n$ and I_n is symmetric we get $I_n = (AA^{-1})^{Tr} = (A^{-1})^{Tr}A^{Tr}$. In exactly the same way it follows from $A^{-1}A = I_n$ that , $A^{Tr}(A^{-1})^{Tr} = I_n$. □

Remark 3.16. As a consequence **Theorem** (3.4.13) anything we can say about the columns of an **invertible matrix** we can also say about the rows as well as many other results. For example, we can prove the following:

Theorem 3.4.14. *Let A be an $n \times n$ matrix. Then the following are equivalent:*

1. *A is* **invertible***;*
2. *A has a* **left inverse***.*

Proof. If A is **invertible** then A^{-1} is a **left inverse** so there is nothing to prove.

Conversely, suppose there is a matrix B such that $BA = I_n$. Taking **transposes** and using **Theorem** (3.3.3) we get that

$$A^{Tr}B^{Tr} = (BA)^{Tr} = I_n^{Tr} = I_n.$$

Therefore A^{Tr} has a **right inverse**. By (3.4.12), A^{Tr} is **invertible**. Since $A = (A^{Tr})^{Tr}$, by **Theorem** (3.4.13), A is **invertible**. □

The following theorem summarizes all the conditions that are equivalent to the **invertibility** of a matrix A.

Theorem 3.4.15. *Let A be an $n \times n$ matrix. Then the following are equivalent:*

1. *A is* **invertible***;*
2. *A has* **reduced echelon form** *the* **identity matrix***, I_n;*
3. *The matrix equation $A\boldsymbol{x} = \mathbf{b}$ has a solution for every vector $\mathbf{b} \in \mathbb{R}^n$.*
4. *The matrix equation $A\boldsymbol{x} = \mathbf{b}$ has at most one solution for each vector $\mathbf{b} \in \mathbb{R}^n$.*
5. *The* **null space** *of A is the* **zero subspace** *of $\mathbb{R}^n$, $null(A) = \{\mathbf{0}_n\}$.*
6. *The* **homogeneous linear system** *with* **coefficient matrix** *A has only the* **trivial solution***.*
7. *A has a* **right inverse***.*
8. *A has a* **left inverse***.*
9. *Each column of A is a* **pivot column** *.*
10. *The sequence of columns of A is* **linearly independent***.*
11. *Every row of A contains a* **pivot position***.*
12. *The sequence of columns of A* **spans** *$\mathbb{R}^n$.*
13. *The sequence of rows of A* **linearly independent***.*
14. *The sequence of rows of A* **spans** *$\mathbb{R}^n$.*
15. *The sequence of columns of A is a* **basis** *of $\mathbb{R}^n$.*
16. *The sequence of rows of A is a* **basis** *of $\mathbb{R}^n$.*

We conclude with one last theorem proving, as promised, the converse to part 1) of **Theorem** (3.4.7):

Theorem 3.4.16. *Let A and B be $n \times n$ matrices and assume that AB is* **invertible***. Then both A and B are invertible.*

Proof. We prove the **contrapositive statement**: If A or B is **non-invertible** then AB is non-invertible. By part 5) of **Theorem**(3.4.15) it suffices to show that the **null space** of AB is not the **zero subspace** of $\mathbb{R}^n$.

Assume first that B is **non-invertible**. Then there exists a non-zero **n-vector**, $\boldsymbol{x}$, such that $B\boldsymbol{x} = \mathbf{0}_n$. But then $(AB)\boldsymbol{x} = A(B\boldsymbol{x}) = A\mathbf{0}_n = \mathbf{0}_n$ and $\boldsymbol{x} \in null(AB)$, whence AB is non-invertible.

We may therefore assume that B is **invertible** and A is **non-invertible**. Then by part 5) of **Theorem**(3.4.15) there exists a non-zero **n-vector**, $\boldsymbol{y}$, such that $A\boldsymbol{y} = \mathbf{0}_n$. Set

$\boldsymbol{x} = B^{-1}\boldsymbol{y}$. Since $\boldsymbol{y} \neq \mathbf{0}_n$ and B^{-1} is **invertible** it follows, again by part 5) of **Theorem** (3.4.15) that $\boldsymbol{x} \neq \mathbf{0}_n$. However, $(AB)\boldsymbol{x} = A(B\boldsymbol{x}) = A(B(B^{-1}\boldsymbol{y})) = A([BB^{-1}]\boldsymbol{y}) = A(I_n\boldsymbol{y}) = A\boldsymbol{y} = \mathbf{0}_n$. Thus, $\boldsymbol{x}$ is in then **null space** of AB and, consequently, AB is **non-invertible**. □

What You Now Can Do

1. Determine if the $n \times n$ matrix A is **invertible**.

2. If an $n \times n$ matrix A is **invertible** compute its **inverse**.

3. If the $n \times n$ matrix A is **invertible** solve a **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$ using the **inverse** of A.

Method (How To Do It)

Method 3.4.1. Determine if the $n \times n$ matrix A is **invertible**.

Use **Gaussian elimination** to obtain an **echelon form** of A. If every row (column) has a **pivot position** (every column is a **pivot column**) then the matrix is **invertible**, otherwise it is **non- invertible**.

Example 3.4.3. Determine if the matrix $\begin{pmatrix} 3 & 2 & 1 \\ 4 & 3 & 3 \\ -1 & 0 & 2 \end{pmatrix}$ is invertible.

Use **Gaussian elimination** to obtain an **echelon form** .

$$\to \begin{pmatrix} 3 & 2 & 1 \\ 0 & \frac{1}{3} & \frac{5}{3} \\ 0 & \frac{2}{3} & \frac{7}{3} \end{pmatrix} \to \begin{pmatrix} 3 & 2 & 1 \\ 0 & \frac{1}{3} & \frac{5}{3} \\ 0 & 0 & -1 \end{pmatrix}$$

Each column is a **pivot column** and consequently the matrix is **invertible**.

Example 3.4.4. Is the matrix

$$\begin{pmatrix} 1 & 2 & 1 & 2 \\ -1 & -3 & 0 & 1 \\ 2 & 0 & 6 & -2 \\ 3 & 4 & 5 & 0 \end{pmatrix} \tag{3.9}$$

invertible?

We apply **Gaussian elimination** to obtain an **echelon form** :

$$\to \begin{pmatrix} 1 & 2 & 1 & 2 \\ 0 & -1 & 1 & 3 \\ 0 & -4 & 4 & -6 \\ 0 & -2 & 2 & -6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & 2 \\ 0 & -1 & 1 & 3 \\ 0 & 0 & 0 & 18 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrix in (3.9) is **non-invertible**.

Method 3.4.2. If an $n \times n$ matrix A is **invertible** compute its **inverse**.

This is done using the **inverse algorithm**.

Example 3.4.5. Determine if the matrix $\begin{pmatrix} 2 & 1 & 1 \\ 3 & 2 & 1 \\ 1 & -1 & 1 \end{pmatrix}$ is **invertible** and, if so, compute the **inverse**.

We augment the matrix by the 3×3 **identity matrix** and use **Gaussian elimination** to obtain an **reduced echelon form** :

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

by switching the first and third rows (this will simplify the calculations by delaying the introduction of fractions).

$$\to \begin{pmatrix} 1 & -1 & 1 & | & 0 & 0 & 1 \\ 0 & 5 & -2 & | & 0 & 1 & -3 \\ 0 & 3 & -1 & | & 1 & 0 & -2 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 1 & | & 0 & 0 & 1 \\ 0 & 5 & -2 & | & 0 & 1 & -3 \\ 0 & 0 & \frac{1}{5} & | & 1 & -\frac{3}{5} & -\frac{1}{5} \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & -1 & 1 & | & 0 & 0 & 1 \\ 0 & 5 & -2 & | & 0 & 1 & -3 \\ 0 & 0 & 1 & | & 5 & -3 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 & | & -5 & 3 & 2 \\ 0 & 5 & 0 & | & 10 & -5 & -5 \\ 0 & 0 & 1 & | & 5 & -3 & -1 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & -1 & 0 & | & -5 & 3 & 2 \\ 0 & 1 & 0 & | & 2 & -1 & -1 \\ 0 & 0 & 1 & | & 5 & -3 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & -3 & 2 & 1 \\ 0 & 1 & 0 & | & 2 & -1 & -1 \\ 0 & 0 & 1 & | & 5 & -3 & -1 \end{pmatrix}$$

The inverse is $\begin{pmatrix} -3 & 2 & 1 \\ 2 & -1 & -1 \\ 5 & -3 & -1 \end{pmatrix}$ (check this).

Example 3.4.6. Determine if the matrix $\begin{pmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & -2 & 1 \\ 1 & 2 & -1 & 2 \\ 1 & 2 & -2 & 2 \end{pmatrix}$ is **invertible** and, if so, compute the **inverse**.

We augment the matrix with I_4 and reduce:

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

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

$$\left(\begin{array}{cccc|cccc} 1 & 1 & 0 & 0 & 1 & 4 & -2 & -2 \\ 0 & 1 & 0 & 0 & -1 & -2 & 4 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \to \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 2 & 6 & -6 & -1 \\ 0 & 1 & 0 & 0 & -1 & -2 & 4 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right)$$

We conclude that the matrix is **invertible** and that its inverse is

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

Method 3.4.3. If the $n \times n$ matrix A is **invertible** solve a **linear system** with **augmented matrix** $[A|\boldsymbol{b}]$ using the **inverse** of A.

Assuming the **coefficient matrix** A of the **linear system** is **invertible** there is a unique solution to the system and it found by multiplying $\boldsymbol{b}$ by A^{-1} on the left hand side: the unique solution is $A^{-1}\boldsymbol{b}$.

Example 3.4.7. Consider the system

$$\begin{array}{rcrcrcr} x & + & 2y & + & 2z & = & 4 \\ 3x & + & 5y & + & 2z & = & -1 \\ 4x & + & 7y & + & 5z & = & 3 \end{array}$$

which can be represented by the **augmented matrix**

$$\left(\begin{array}{ccc|c} 1 & 2 & 2 & 4 \\ 3 & 5 & 2 & -1 \\ 4 & 7 & 5 & 3 \end{array}\right)$$

or as a matrix equation

$$\begin{pmatrix} 1 & 2 & 2 \\ 3 & 5 & 2 \\ 4 & 7 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix}$$

The **coefficient matrix** of the **linear system** is **invertible with inverse**

$$\begin{pmatrix} -11 & -4 & 6 \\ 7 & 3 & -4 \\ -1 & -1 & 1 \end{pmatrix}$$

There is a unique solution to the system and it is

$$\begin{pmatrix} -11 & -4 & 6 \\ 7 & 3 & -4 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} -22 \\ 13 \\ 0 \end{pmatrix}.$$

Exercises

For each of the matrices in exercises 1 - 14 determine if it has an inverse and, if so, compute the inverse. See either **Theorem** (3.4.3) or **Method** (3.4.2).

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

2. $\begin{pmatrix} 2 & 5 \\ 3 & 7 \end{pmatrix}$

3. $\begin{pmatrix} 3 & -1 \\ -3 & 4 \end{pmatrix}$

4. $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$

5. $\begin{pmatrix} 1 & -3 & -1 \\ -2 & 5 & 1 \\ 3 & 7 & 12 \end{pmatrix}$

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

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

8. $\begin{pmatrix} 4 & 5 & 2 \\ 1 & 1 & 1 \\ 2 & 3 & 1 \end{pmatrix}$

9. $\begin{pmatrix} 2 & 1 & 1 \\ 7 & 3 & 2 \\ 3 & 2 & 4 \end{pmatrix}$

10. $\begin{pmatrix} 1 & 1 & 2 & 0 \\ 2 & 1 & 0 & 4 \\ 2 & 0 & -3 & 8 \\ 2 & 5 & 11 & -12 \end{pmatrix}$

11. $\begin{pmatrix} 2 & 5 & 3 & 4 \\ 3 & 7 & 3 & 7 \\ 1 & 3 & 2 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix}$

12. $\begin{pmatrix} 1 & 2 & 0 & 4 \\ 1 & 1 & 3 & 6 \\ 2 & 3 & 0 & 8 \\ 2 & 4 & 1 & 9 \end{pmatrix}$

13. $\begin{pmatrix} 4 & -7 & 1 & 5 \\ -3 & 5 & -1 & -3 \\ 1 & -2 & 1 & -1 \\ 2 & -4 & 0 & 5 \end{pmatrix}$

14. $\begin{pmatrix} 4 & -7 & 1 & 5 \\ -3 & 5 & -1 & -3 \\ 1 & -2 & -1 & -1 \\ 2 & -4 & 0 & 5 \end{pmatrix}$

15. Let $A = \begin{pmatrix} 1 & -1 & 1 \\ 2 & -1 & 3 \\ -2 & 3 & 0 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 7 & 3 \\ 1 & 5 & 2 \end{pmatrix}$. Demonstrate by computing that $(AB)^{-1} = B^{-1}A^{-1}$.

In exercises 16-19 determine if the **coefficient matrix** of the given **linear system** is **invertible**. If it is use the inverse to find the unique **solution** to the system.

16.

$$\begin{array}{ccccccc} x & + & 2y & + & 5z & = & 2 \\ 2x & + & 3y & + & 7z & = & 0 \\ 4x & - & y & - & 6z & = & -3 \end{array}$$

17.

$$\begin{array}{ccccccc} x & + & y & + & 2z & = & -1 \\ 3x & + & 2y & + & 4z & = & -3 \\ 2x & + & 5y & + & 9z & = & 2 \end{array}$$

18.

$$\begin{array}{ccccccc} 2x & + & y & & & = & 1 \\ 5x & + & 3y & + & z & = & 1 \\ x & + & 2y & + & 3z & = & 1 \end{array}$$

19.

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

In exercises 20-24 answer true or false and give and an explanation.

20. If A and B are 2×2 **invertible matrices** then $A + B$ is invertible.

21. If A and B are 2×2 **non-invertible matrices** then $A + B$ is non-invertible.

22. If A is an **invertible** $n \times n$ matrix then $null(A) = null(A^{-1})$.

23. If A is a 3×3 matrix and $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is in $null(A)$ then $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in null(A^{-1})$.

24. If A is an **invertible** $n \times n$ matrix then the **null space of A** is equal to the null space of the **transpose** of A, $null(A) = null(A^{Tr})$.

Challenge Exercises (Problems)

1. Let A be an $m \times n-$matrix. Show that the following are equivalent:

a) There exists an $n \times m-$matrix B such that BA is **invertible**.
b) There exists an $n \times m-$matrix C such that $CA = I_n$.

2. Let A be an $m \times n-$matrix and assume that there exists an $n \times m-$matrix C such that $CA = I_n$.

a) Prove that the **null space of** A is the **zero subspace**, $null(A) = \{\mathbf{0}_n\}$.
b) Prove that the sequence of columns of A is **linearly independent**.

3. Let A be an $m \times n$ matrix and assume that the sequence of columns of A is **linearly independent**.

a) Prove that the **transpose** of A, A^{Tr}, has n **pivot positions**.

b) Prove that the sequence of columns of A^{Tr} **spans** $\mathbb{R}^n$.

c) Prove that there is an $n \times m-$matrix C such that $CA = I_n$.

4. a) Verify that the vectors $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$ are both **solutions** to the **linear system**

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

b) Explain why the **coefficient matrix**, $\begin{pmatrix} 1 & 2 & -1 \\ 2 & 3 & -1 \\ 2 & 5 & -3 \end{pmatrix}$, of the **linear system** is **non-invertible**. See **Theorem** (3.4.15).

5. a) Verify that the **linear system** with augmented matrix $\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 2 & -1 & 1 \\ 3 & 2 & 5 & 2 \end{array}\right)$ is **inconsistent**.

b) Explain why the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 3 & 2 & 5 \end{pmatrix}$ is **non-invertible**. See **Theorem** (3.4.15).

c) Explain why the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 3 & 2 & 2 \end{pmatrix}$ is **invertible**. See **Theorem** (3.4.15).

6. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ -5 \end{pmatrix}$.

a) Prove that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} : x + y + z = 0 \}$.

b) Let $A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 1 \\ -3 & -5 & -4 \end{pmatrix}$. Explain why $null(A) = \{\mathbf{0}_3\}$. See **Theorem** (2.4.11) and **Theorem** (3.4.15)

In question 7 and 8 let A be a $3 \times 3-$matrix with columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$, where the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is **linearly independent**.

7. Assume the **linear system** with **augmented matrix** $[A|\boldsymbol{v}_3]$ has a unique **solution**. Explain why A is **invertible**. See **Theorem** (3.4.15).

8. Now assume that the **linear system** with **augmented matrix** $[A| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}]$ is **inconsistent**. Explain why $\boldsymbol{v}_3$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2)$. See **Theorem** (3.4.15).

9. Assume that A and Q are **invertible** $n \times n$ matrices and set $B = Q^{-1}AQ$. Prove that B is invertible and that $B^{-1} = Q^{-1}A^{-1}Q$.

10. Assume that A, B are $n \times n$ matrices and that $A^2 = B^2 = (AB)^2 = I_n$. Prove that $AB = BA$.

Quiz Solutions

1. The **reduced echelon form** of A is the **3×3 identity matrix, I_3**. Therefore very row has a **pivot position** and the sequence of columns of A **spans** $\mathbb{R}^3$.

Not right, see **Method** (2.3.2).

2. By 1) every column of A is a **pivot column** and consequently the sequence of columns of A is **linearly independent**.

Not right, see **Method** (2.4.1).

3. Since the sequence of columns of A **spans** $\mathbb{R}^3$, in particular, each of the **standard**

basis vectors, e_i^3 of $\mathbb{R}^3$ is a **linear combination** of the sequence of columns of A. This means there are vectors $\boldsymbol{v}_i, i = 1, 2, 3$ such that the **product** of A with $\boldsymbol{v}_i$ is $\boldsymbol{e}_i^3$, that is, $A\boldsymbol{v}_i = \boldsymbol{e}_i^3$.

Not right, see **Theorem** (3.2.2).

4. $\boldsymbol{v}_1 = \begin{pmatrix} -18 \\ 8 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} -25 \\ 11 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} -23 \\ 10 \\ 1 \end{pmatrix}$

Not right, see **Method** (2.2.2)

5. $AB = (A\boldsymbol{v}_1 A\boldsymbol{v}_2 A\boldsymbol{v}_3) = (\boldsymbol{e}_1^3\ \boldsymbol{e}_2^3\ \boldsymbol{e}_3^3) = I_3$.

Not right, see **Definition** (3.18).

6. $BA = I_3$.

Not right, **Definition** (3.18).

3.5. Elementary Matrices

We introduce the concept of an elementary matrix and that row operations can be realized via left multiplication by elementary matrices. We also develop links between invertibility and elementary matrices.

Am I Ready for This Material

Extensive use is made of the following concepts and procedures in this section and they need to be mastered to fully understand the new ideas and methods introduced here:

linear system

consistent and inconsistent linear system

matrix

elementary row operation

matrix in echelon form

matrix in reduced echelon form

pivot positions of a matrix

an echelon form of a matrix

the reduced echelon form of a matrix

pivot columns of a matrix

row equivalence of matrices

linear combination

span of a sequence of vectors

linearly dependent sequence of vectors

linearly independent sequence of vectors

null space of a matrix

inverse of an invertible matrix

transpose of a matrix

Important procedures that you should have mastered are:
Gaussian elimination
matrix inversion algorithm

Quiz

In 1 - 3 determine if the given sequence of vectors is **linearly independent**.

1. $\left(\begin{pmatrix}1\\1\\2\end{pmatrix}, \begin{pmatrix}3\\2\\4\end{pmatrix}, \begin{pmatrix}2\\3\\6\end{pmatrix}\right)$.

2. $\left(\begin{pmatrix}2\\3\\1\end{pmatrix}, \begin{pmatrix}3\\5\\2\end{pmatrix}, \begin{pmatrix}3\\4\\4\end{pmatrix}\right)$.

3. $\left(\begin{pmatrix}1\\-2\\3\end{pmatrix}, \begin{pmatrix}1\\-3\\2\end{pmatrix}, \begin{pmatrix}1\\-1\\5\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right)$.

In 4 - 6 determine if the given sequence of vectors **span** the appropriate $\mathbb{R}^n$.

4. $\left(\begin{pmatrix}1\\-2\\3\\1\end{pmatrix}, \begin{pmatrix}1\\-3\\2\\2\end{pmatrix}, \begin{pmatrix}1\\-1\\5\\3\end{pmatrix}\right)$.

5. $\left(\begin{pmatrix}4\\3\\2\\1\end{pmatrix}, \begin{pmatrix}3\\2\\1\\0\end{pmatrix}, \begin{pmatrix}2\\1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\0\\0\end{pmatrix}\right)$.

6. $\left(\begin{pmatrix}1\\2\\-1\\-2\end{pmatrix}, \begin{pmatrix}2\\3\\-1\\-4\end{pmatrix}, \begin{pmatrix}-1\\3\\4\\-6\end{pmatrix}, \begin{pmatrix}1\\1\\1\\-3\end{pmatrix}\right)$.

In 7 - 8 determine if the given matrix is **invertible** and, if so, find its **inverse**.

7. $\begin{pmatrix}1 & 2 & 2\\3 & 5 & 7\\1 & -1 & 5\end{pmatrix}$.

8. $\begin{pmatrix}1 & 2 & 2\\3 & 5 & 7\\1 & -1 & 4\end{pmatrix}$.

Quiz Solutions

New Concepts

The main new concept in this section is that of an **elementary matrix**. Such a matrix corresponds to an **elementary row operation** . Therefore there are three types: exchange, scaling and elimination.

Theory (Why It Works)

We have seen the importance of **elementary row operations** since these are the main tool we have for such things as finding the **null space of a matrix**, determining if a sequence of vectors **spans** $\mathbb{R}^n$, and so on.

We have also seen the importance of **invertible matrices** to solving certain types of **linear systems**. In the present section we relate **elementary row operations** to invertible matrices and prove a number of very powerful results.

Definition 3.24. An ***elementary matrix*** is a **square matrix** obtained from the **identity matrix** of the same size by the application of an **elementary row operation**. There are three types, one of each of the three elementary row operations: exchange, scaling, and elimination.

These are perhaps best understood by looking at examples.

Example 3.5.1. Elementary matrices obtained from exchange

Exchanging the first and second rows of the 2×2 **identity matrix** yields

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

Exchanging the second and third rows of the 3×3 **identity matrix** gives the following **elementary matrix**:

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

By exchanging the second and fourth rows of the 4×4 **identity matrix** we get:

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

Notation

We will denote the matrix obtained from the $n \times n$ **identity matrix** , I_n, by exchanging the i^{th} and j^{th} rows by P_{ij}^n (the P stands for "permutation"). So, in **Example** (3.5.1) the first matrix is P_{12}^2, the second is P_{23}^3 and the third is P_{24}^4.

Example 3.5.2. Elementary matrices obtained by scaling

Scaling the second row of the 2×2 **identity matrix** by $-\frac{2}{5}$ gives the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -\frac{2}{5} \end{pmatrix}$$

If we multiply the third row of the 3×3 **identity matrix** by 4 we obtain the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

As a final example, multiplying the first row of the 4×4 **identity matrix** by -7 gives

$$\begin{pmatrix} -7 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Notation

We denote by $D_i^n(c)$ the matrix obtained from the $n \times n$ **identity matrix** , I_n, by multiplying the i^{th} row by c. Thus, in **Example** (3.5.2) the first matrix is $D_2^2(-\frac{2}{5})$, the second is $D_3^3(4)$, and the last one is $D_1^4(-7)$.

Example 3.5.3. Elementary matrices obtained by elimination

Adding $-\frac{2}{3}$ times the first row of the 2×2 **identity matrix** to the second row yields the matrix

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

Adding -3 times the third row of the 3×3 **identity matrix** to the first row produces the matrix

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

Finally adding $\frac{3}{8}$ times the fourth row to the second row of the 4×4 **identity matrix** gives the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{3}{8} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Notation

We shall denote the matrix obtained from the $n \times n$ **identity matrix** , I_n, by adding c times the i^{th} row to the j^{th} row by $T_{ij}^n(c)$. Thus, the first matrix appearing in **Example** (3.5.3) is $T_{12}^2(-\frac{2}{3})$. The second one is denoted by $T_{31}^3(-3)$. The final example is $T_{42}^4(\frac{3}{8})$.

Remark 3.17. An important property of **elementary matrices** is that, when a matrix A is multiplied on the left by an elementary matrix E the resulting product is equal to the matrix obtained by performing the corresponding **elementary row operation** on A.

Example 3.5.4. Multiplying the matrix

$$\begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix}$$

by the **exchange matrix** P_{13}^3 gives the following:

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix} = \begin{pmatrix} 7 & -8 & 9 \\ -4 & 5 & -6 \\ 1 & -2 & 3 \end{pmatrix}.$$

Multiplying the matrix

$$\begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix}$$

by the **elimination matrix** $T_{23}^3(-2)$ gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 15 & -18 & 21 \end{pmatrix}$$

Finally, multiplying the matrix

$$\begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix}$$

by the **scaling matrix** $D_2^3(3)$ yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 3 \\ -12 & 15 & -18 \\ 7 & -8 & 9 \end{pmatrix}.$$

Since the effect of an **elementary row operation** can be reversed by applying an elementary row operation of the same type it follows that an **elementary matrix** is **invertible** and that the **inverse** is an elementary matrix of the same type. This is summarized in the following result:

Theorem 3.5.1. *Every* **elementary matrix** *is* **invertible**. *Moreover the* **inverse** *of an* **elementary matrix** *is an elementary matrix. In fact,*

1. $(P_{ij}^n)^{-1} = P_{ij}^n$
2. $T_{ij}^n(c)^{-1} = T_{ij}^n(-c)$
3. $D_i^n(c)^{-1} = D_i^n(\frac{1}{c})$.

Recall that two matrices, A, B, are said to be **row equivalent** if they have the same **reduced echelon form** and this occurs if and only if there is a sequence of **elementary row operations** that transform A into B. We write $A \sim B$ when A and B are **row equivalent**.

Because each **elementary row operation** performed on a matrix A can be achieved by multiplying A on the left by the corresponding **elementary matrix** the following theorem holds:

Theorem 3.5.2. *Two $n \times n$ matrices A and B are* **row equivalent** *if and only if there is a sequence $(E_1, E_2, \ldots, E_k)$ of* **elementary matrices** *such that $B = E_k E_{k-1} \ldots E_2 E_1 A$.*

As a special case we can apply the **Theorem** (3.5.2) to a matrix A and its **reduced echelon form**. Thus, suppose A is a matrix and U its **reduced echelon form**. Then there is a sequence $(E_1, E_2, \ldots, E_k)$ of **elementary matrices**

$$U = E_k E_{k-1} \ldots E_2 E_1 A.$$

Since each E_i is invertible, it then also follows that $A = E_1^{-1} E_2^{-1} \ldots E_{k-1}^{-1} E_k^{-1} U$.

Example 3.5.5. Apply **Gaussian elimination** to the matrix $\begin{pmatrix} 0 & 3 & 8 \\ 2 & -1 & -2 \\ 1 & -2 & 2 \end{pmatrix}$ to obtain its **reduced echelon form** U and then write U as a **product** of A and **elementary matrices**.

The steps of the **Gaussian elimination** are as follows:

$$\to \begin{pmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 0 & 3 & 8 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 2 \\ 0 & 3 & -6 \\ 0 & 3 & 8 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 2 \\ 0 & 1 & -2 \\ 0 & 3 & 8 \end{pmatrix} \to$$
$$\begin{pmatrix} 1 & -2 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 14 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to$$
$$\begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The **elementary matrices** used, in the order in which they are multiplied (from left to right), are

$$P_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, T_{12}(-2) = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$D_2(\frac{1}{3}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}, T_{23}(-3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix},$$

$$D_3(\frac{1}{14}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{14} \end{pmatrix}, T_{32}(2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix},$$

$$T_{31}(-2) = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, T_{21}(2) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus,

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{14} \end{pmatrix}$$

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

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

By taking inverses, we obtain the following expression for A as a product of elementary matrices:

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

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{14} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} =$$

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

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 14 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So, we have expressed A as a **product** of **elementary matrices**.

The following is essentially an observation

Lemma 3.5.3. *The* **reduced echelon form** *R of an $n \times n$ matrix is either the $n \times n$* **identity matrix** *, I_n, or has a* **zero row**.

Proof. Since R is a square matrix, if each row of R has a **pivot position** then R is the $n \times n$ **identity matrix** , I_n. On the other hand if some row does not have a **pivot position** then it is a **zero row**. $\square$

When we combine **Lemma** (3.5.3) with the implications of **Theorem** (3.5.2) we get the following result:

Theorem 3.5.4. *The following are equivalent:*

1. *The $n \times n$ matrix A is* **invertible**;

2. *A is* **row equivalent** *to the $n \times n$* **identity matrix** *, I_n;*

3. *A is a* **product** *of* **elementary matrices**.

Proof. 1) implies 2) Assume that A is **invertible**. Then by 11) of **Theorem** (3.4.15) each row of A has a **pivot position** and so by the **Theorem** (3.5.4) A is **row equivalent** the $n \times n$ **identity matrix** , I_n.

2) implies 3) Suppose A is **row equivalent** to the $n \times n$ **identity matrix** . Then there is a sequence of **elementary matrices** $(E_1, E_2, \ldots, E_k)$ such that

$$I_n = E_k E_{k-1} \ldots E_2 E_1 A.$$

But it then if follows that

$$A = E_1^{-1} E_2^{-1} \ldots E_{k-1}{-1} E_k^{-1} I_n = E_1^{-1} E_2^{-1} \ldots E_{k-1}^{-1} E_k^{-1}.$$

Since the **inverse** of an **elementary matrix** is an elementary matrix by **Theorem** (3.5.1) it follows that A is a **product** of **elementary matrices**.

3) implies 1) Assume A is a **product** of **elementary matrices**. Elementary matrices are **invertible** by **Theorem** (3.5.1). Additionally, as indicated in **Remark** (3.12) the **product** of finitely many **invertible matrices** is invertible. Therefore, A is invertible. $\square$

Example (3.5.5) suggests another explanation for why the **matrix inversion algorithm** works. Recall, for an $n \times n$ matrix A we form the augmented matrix $[A|I_n]$ and then we apply **Gaussian elimination** until we either get a **zero row** on the left hand side (and conclude that A is **non-invertible**) or the $n \times n$ **identity matrix** , I_n. We claimed that the matrix obtained on the right hand side is A^{-1}.

Let us suppose that A is **invertible**. Then the **reduced echelon form** of A is I_n. We have seen that this implies that there is a sequence $(E_1, E_2, \ldots, E_k)$ of **elementary matrices** such that $E_k E_{k-1} \ldots E_2 E_1 A = I_n$.

When we perform the sequence of **elementary row operation** corresponding to these **elementary matrices** in the given order to $[A|I_n]$ we obtain

$$[E_k E_{k-1} \ldots E_2 E_1 A \mid E_k E_{k-1} \ldots E_2 E_1 I_n] = [E_k E_{k-1} \ldots E_2 E_1 A | E_k E_{k-1} \ldots E_2 E_1].$$

However, since $(E_k E_{k-1} \ldots E_2 E_1)A = I_n$ it follows that $E_k E_{k-1} \ldots E_2 E_1 = A^{-1}$.

Example 3.5.6. Write the **inverse** of the matrix $A = \begin{pmatrix} 2 & 5 \\ 3 & 7 \end{pmatrix}$ as a **product** of **elementary matrices**.

Note that for a 2×2 matrix we know how to find the **inverse** directly by **Theorem** (3.4.3), and we can use this as a check. For the matrix A we have $A^{-1} = \frac{1}{-1}\begin{pmatrix} 7 & -5 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} -7 & 5 \\ 3 & -2 \end{pmatrix}$.

We express the matrix $\begin{pmatrix} 2 & 5 \\ 3 & 7 \end{pmatrix}$ as a **product** of **elementary matrices** in two different ways. Straightforward application of **Gaussian elimination** yields the following:

$$\begin{pmatrix} 2 & 5 \\ 3 & 7 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{5}{2} \\ 3 & 7 \end{pmatrix} \to \begin{pmatrix} 1 & \frac{5}{2} \\ 0 & -\frac{1}{2} \end{pmatrix} \to \begin{pmatrix} 1 & \frac{5}{2} \\ 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

From this we conclude that

$$A^{-1} = T_{21}(-\frac{5}{2})D_2(-2)T_{12}(-3)D_1(\frac{1}{2}) =$$

$$\begin{pmatrix} 1 & -\frac{5}{2} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -7 & 5 \\ 3 & -2 \end{pmatrix}.$$

Another strategy is to get a 1 in the first column by subtracting and avoid the introduction of fractions. We get the following reduction:

$$\begin{pmatrix} 2 & 5 \\ 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \to \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now we get

$$A^{-1} = T_{21}(-2)T_{12}(-2)P_{12}T_{12}(-1) =$$

$$\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -7 & 5 \\ 3 & -2 \end{pmatrix}.$$

Check that these products are equal to A^{-1}.

Example 3.5.7. Express the **inverse** of the matrix B = $\begin{pmatrix} 3 & 4 & -6 \\ 1 & 2 & -3 \\ 2 & 3 & -4 \end{pmatrix}$ as a **product** of **elementary matrices**.

The **Gaussian elimination** is as follows:

$$\begin{pmatrix} 3 & 4 & -6 \\ 1 & 2 & -3 \\ 2 & 3 & -4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -3 \\ 3 & 4 & -6 \\ 2 & 3 & -4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -3 \\ 0 & -2 & 3 \\ 2 & 3 & -4 \end{pmatrix} \to$$
$$\begin{pmatrix} 1 & 2 & -3 \\ 0 & -2 & 3 \\ 0 & -1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -3 \\ 0 & -1 & 2 \\ 0 & -2 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & -2 \\ 0 & -2 & 3 \end{pmatrix} \to$$
$$\begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to$$
$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So,

$$A^{-1} = T_{21}(-2)T_{31}(3)T_{32}(2)D_3(-1)T_{23}(2)D_2(-1)P_{23}T_{13}(-2)T_{12}(-3)P_{12}.$$

Since every **invertible matrix** is a **product** of **elementary matrices** we get another criterion for **row equivalence** of two matrices.

Theorem 3.5.5. *Let A and B be $m \times n$ matrices. Then A and B are* **row equivalent** *if and only if there is an* **invertible** *$m \times m$ matrix Q such that $B = QA$.*

Proof. We leave the proof as a **Challenge Exercise**

What You Can Now Do

1. Write down the **elementary matrix** corresponding to a given **elementary row operation**.

2. Express a **square matrix** A as a **product** of **elementary matrices** and the **reduced echelon form** of A.

3. Express a **non-singular matrix** and/or its **inverse** as a **product** of **elementary matrices**.

Method (How To Do It)

Method 3.5.1. Write down the **elementary matrix** corresponding to a given **elementary row operation**.

Apply the **elementary row operation** to the **identity matrix**.

Example 3.5.8. Write down the **elementary matrices** which correspond to the following **elementary row operations**.

a) On matrices with four rows: $R_2 \leftrightarrow R_3$.

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

This is the matrix P_{23}^4 (see **notation for a exchange matrix**).

b) On matrices with three rows: $R_1 \leftrightarrow R_3$.

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

This is the matrix P_{13}^3 (see **notation for a exchange matrix**).

c) On matrices with three rows: $R_3 \to (-\frac{3}{5})R_3$.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{3}{5} \end{pmatrix}$$

This is the matrix $D_3^3(-\frac{3}{5})$ (see **notation for a scaling matrix**).

d) On matrices with four rows: $R_2 \to \frac{5}{4}R_2$.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{5}{4} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The notation for this matrix is $D_2^4(\frac{5}{4})$ (see **notation for a scaling matrix**).

e) On matrices with three rows: Add $R_2 \to R_2 + (-\frac{4}{3})R_3$.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{4}{3} \\ 0 & 0 & 1 \end{pmatrix}$$

This matrix is denoted by $T_{32}^3(-\frac{4}{3})$ (see **notation for an elimination matrix**).

f) For matrices with four rows: $R_3 \to (-\frac{1}{2})R_1 + R_3$.

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

This is the matrix $T_{13}^4(-\frac{1}{2})$ (see **notation for an elimination matrix**).

Method 3.5.2. Express a **square matrix** A as a **product** of **elementary matrices** and the **reduced echelon form** of A.

To do this keep track of the **elementary row operation** when applying **Gaussian elimination** to a matrix to obtain its **reduced echelon form**. Write down the corresponding **elementary matrices**. Then multiply the **reduced echelon form** obtained by the **inverses** of the **elementary matrices** in the reverse order.

Example 3.5.9. Write the matrix $A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 4 & 4 \\ 1 & 2 & 2 \end{pmatrix}$ as a **product** of **elementary matrices** and the **reduced echelon form** of A.

We apply **Gaussian elimination** :

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

$$\begin{pmatrix} 1 & 3 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} = R.$$

To accomplish this we multiplied on the left by the following **elementary matrices** (in the given order):

$$T_{12}^3(-2), T_{13}^3(-1), T_{23}^3(-\frac{1}{2}), D_2^3(-\frac{1}{2}), T_{21}^3(-3).$$

This means that

$$T_{21}^3(-3)D_2^3(-\frac{1}{2})T_{23}^3(-\frac{1}{2})T_{13}^3(-1)T_{12}^3(-2)A = R.$$

Then

$$A = T_{12}^3(-2)^{-1}T_{13}^3(-1)^{-1}T_{23}^3(-\frac{1}{2})^{-1}D_2^3(-\frac{1}{2})^{-1}T_{21}^3(-3)^{-1}R =$$

$$T_{12}^3(2)T_{13}^3(1)T_{23}^3(\frac{1}{2})D_2^3(-2)T_{21}^3(3)R.$$

Method 3.5.3. Express a **non-singular matrix** A and/or its **inverse** as a **product** of **elementary matrices**.

Apply **Gaussian elimination** , keeping track of the **elementary row operations**. For each **elementary row operation** write down the corredsponding **elementary matrix** in the given order. The **product** of these **elementary matrices** will be the **inverse** of the matrix A.

By taking the **inverses** of the **elementary matrices** and multiplying them in reverse order we obtain A as a **product** of **elementary matrices**.

Example 3.5.10. If possible, express the following matrix and its **inverse** as a **product** of **elementary matrices**: $B = \begin{pmatrix} -1 & 2 & -1 \\ 3 & -5 & 4 \\ 2 & 0 & 6 \end{pmatrix}$.

We apply **Gaussian elimination** , keeping track of the **elementary row operations**:

$$\to \begin{pmatrix} 1 & -2 & 1 \\ 3 & -5 & 4 \\ 2 & 0 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 1 \\ 2 & 0 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 1 \\ 0 & 4 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Since the **reduced echelon form** of this matrix has a **zero row** it is not **invertible**.

Example 3.5.11. If possible, express the following matrix and its inverse as a **product** of **elementary matrices**:

$$C = \begin{pmatrix} 1 & 3 & 2 & -1 \\ 0 & 4 & -1 & 2 \\ -3 & 4 & 4 & -6 \end{pmatrix}$$

This matrix is not a **square matrix** and consequently cannot be **invertible** or written as a **product** of **elementary matrices**.

Example 3.5.12. If possible, express the following matrix and its inverse as a **product** of **elementary matrices**:

$$D = \begin{pmatrix} 1 & -2 & 1 \\ -4 & 7 & -5 \\ 2 & -3 & 5 \end{pmatrix}$$

We apply **Gaussian elimination** , keeping track of the **elementary row operations**:

$$\to \begin{pmatrix} 1 & -2 & 1 \\ 0 & -1 & -1 \\ 2 & -3 & 5 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 1 \\ 0 & -1 & -1 \\ 0 & 1 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 3 \end{pmatrix} \to$$

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

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

The **elementary matrices** corresponding to the **elementary row operations** , in the order in which they occured, are:

$$T_{12}^3(4), T_{13}^3(-2), D_2^3(-1), T_{23}^3(-1), D_3^3(\frac{1}{2}), T_{31}^3(-1), T_{32}^3(-1), T_{21}^3(2).$$

Thus,

$$T_{21}^3(2)T_{32}^3(-1)T_{31}^3(-1)D_3^3(\frac{1}{2})T_{23}^3(-1)D_2^3(-1)T_{13}^3(-2)T_{12}^3(4)D = I_3$$

Consequently,

$$D^{-1} = T_{21}^3(2)T_{32}^3(-1)T_{31}^3(-1)D_3^3(\frac{1}{2})T_{23}^3(-1)D_2^3(-1)T_{13}^3(-2)T_{12}^3(4)$$

In order to write D as a product of **elementary matrices**, we **invert** each of these and multiply them in the reverse order.:

$$D = T_{12}^3(-4)T_{13}^3(2)D_2^3(-1)T_{23}(1)D_3^3(2)T_{31}^3(1)T_{32}^3(1)T_{21}^3(-2).$$

Exercises

In exercises 1 - 7 write down the **elementary matrix** which corresponds to the **elementary row operation** as applied to the given matrix. See **Method** (3.5.1).

1. $R_2 \to (-2)R_1 + R_2 : \begin{pmatrix} 1 & 0 & -1 \\ 2 & 4 & 3 \\ 5 & -1 & 2 \\ 0 & 1 & 4 \end{pmatrix}.$

2. $R_3 \to (-3)R_1 + R_3 : \begin{pmatrix} 1 & -1 & 2 & -1 \\ -2 & 2 & -5 & 3 \\ 3 & -2 & -5 & 4 \end{pmatrix}.$

3. $R_1 \to R_1 + (-1)R_3 : \begin{pmatrix} 3 & 4 & 2 & 3 \\ 4 & -1 & 0 & 2 \\ 2 & 5 & 1 & -1 \end{pmatrix}.$

4. $R_3 \to \frac{1}{2}R_3 : \begin{pmatrix} 1 & 3 & -2 & 1 \\ 3 & -2 & 1 & 4 \\ 2 & -4 & 6 & 0 \end{pmatrix}.$

5. $R_2 \to (-1)R_2 : \begin{pmatrix} 3 & -1 & 2 \\ -2 & -1 & -3 \end{pmatrix}.$

6. $R_1 \leftrightarrow R_3 : \begin{pmatrix} 4 & 0 & 1 & 2 \\ -1 & 1 & 1 & 3 \\ 1 & 0 & -1 & 0 \end{pmatrix}.$

7. $R_2 \leftrightarrow R_4 : \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ -1 & -5 \\ 0 & 1 \end{pmatrix}.$

In 8 - 11 express the **square matrix** A as the **product** of **elementary matrices** and the **reduced echelon form** of A. See **Method** (3.5.2).

8. $\begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$.

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

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

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

In exercises 12 - 14 express the **square matrix** A and its **inverse** as a **product** of **elementary matrices**. See **Method** (3.5.3).

12. $\begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}$.

13. $\begin{pmatrix} 3 & 5 \\ 5 & 9 \end{pmatrix}$.

14. $\begin{pmatrix} 1 & 1 & -2 \\ 2 & 3 & -5 \\ -1 & -1 & 3 \end{pmatrix}$.

In exercises 15 - 18 answer true or false and give an explanation.

15. If E is an $m \times m$ **elementary matrix**, A is an $m \times n$ matrix and $EA = \mathbf{0}_{m\times n}$ then $A = \mathbf{0}_{m\times n}$.

16. $T_{ij}^n(c)T_{ij}^n(d) = T_{ij}^n(c+d)$.

17. If the $n \times n$ matrix A is a **product** of **elementary matrices** then the **null space** of A is the $n \times n$ **zero subspace**, $null(A) = \{\mathbf{0}_n\}$.

18. If the $n \times n$ matrix A is a **product** of **elementary matrices** then the **transpose** of A, A^{Tr}, is a **product** of **elementary matrices**.

Challenge Exercises (Problems)

1. Prove that two $m \times n$ matrices A and B are **row equivalent** if and only if there exists an **invertible** $m \times m$ matrix Q such that $B = QA$.

2. Assume $1 < j$ and prove that $P_{2j}P_{12}P_{2j} = P_{1j}$.

3. Assume $1 < j$ and prove that $P_{2j}T_{12}(c)P_{2j} = T_{1j}(c)$.

Quiz Solutions

1. No. Computational explanation: After applying **Gaussian elimination** to the matrix $\begin{pmatrix} 1 & 3 & 2 \\ 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}$ we obtain the **reduced echelon form** :

$\begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. The last column is not a **pivot column** and consequently the sequence is not linearly independent.

Not right, see **Method** (2.4.1).

Elegant explanation: The **transpose** of the matrix is $\begin{pmatrix} 1 & 1 & 2 \\ 3 & 2 & 4 \\ 2 & 3 & 6 \end{pmatrix}$. Observe that the third column is twice the second column and therefore the sequence of columns of A^{Tr} is linearly dependent. Since A^{Tr} is a **square matrix** we conclude that A^{Tr} is **non-invertible**. Then A is **non-invertible** by **Theorem** (3.4.13). Then by **Theorem** (3.4.15) we conclude that the sequence of columns of A is linearly dependent.

2. Yes. Applying **Gaussian elimination** to the matrix $\begin{pmatrix} 2 & 3 & 3 \\ 3 & 5 & 4 \\ 1 & 2 & 4 \end{pmatrix}$ we obtain the **reduced echelon form** which is the 3×3 **identity matrix** , I_3. Thus, every column is a **pivot column** and we conclude that the sequence is **linearly independent**.

Not right, see **Theorem** (2.4.1).

3. No. Computation is not required since a sequence of four vectors in $\mathbb{R}^3$ cannot be **linearly independent**.

Not right, see **Theorem** (2.4.2).

4. No. Computation is not required since a sequence of three vectors cannot **span** $\mathbb{R}^4$.

Not right, see **Theorem** (2.3.2).

5. Yes. When we apply **Gaussian elimination** to the matrix $\begin{pmatrix} 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$ we obtain the 4×4 **identity matrix** , I_4. Therefore every row contains a **pivot position** and consequently the sequence **spans** $\mathbb{R}^4$.

Not right, see **Method** (2.3.2).

6. No. Computational explanation: When we apply **Gaussian elimination** to the matrix $\begin{pmatrix} 1 & 2 & -1 & -1 \\ 2 & 3 & 3 & 1 \\ -1 & -1 & 4 & 1 \\ -2 & -4 & -6 & -3 \end{pmatrix}$ we obtain the **reduced echelon form** , $\begin{pmatrix} 1 & 0 & 0 & \frac{15}{8} \\ 0 & 1 & 0 & -\frac{3}{8} \\ 0 & 0 & 1 & \frac{1}{8} \\ 0 & 0 & 0 & 0 \end{pmatrix}$. There is a **zero row** and therefore the sequence does not **span** $\mathbb{R}^4$. Not right, see **Method** (2.3.2).

Elegant explanation. Each of the columns belongs to the **subspace** $S = \{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} | x_1 + x_2 + x_3 + x_4 = 0\}$. Therefore the **subspace** **spanned** by the sequence is contained in S and therefore is not equal to $\mathbb{R}^4$.

7. There is no **inverse**, when the **matrix inversion algorithm** is performed we obtain a **zero row**. Not right, see **Method** (3.4.1)

8. The matrix is **invertible** as can be verified using **Method** (3.4.1). Applying **Method** (3.4.2) we determine that the matrix $\begin{pmatrix} 27 & -10 & 4 \\ -5 & 2 & -1 \\ -8 & 3 & -1 \end{pmatrix}$ is its **inverse**.

3.6. The LU Factorization

In this section we introduce the concept of a factorization of a matrix. One such factorization is the $LU-$decomcomposition which is used in applications where several **linear systems** with the same **coefficient matrix** have to be solved.

Am I Ready for This Material

Extensive use is made of the following concepts and procedures in this section and they need to be mastered to fully understand the new ideas and methods introduced here:

linear system

inhomogeneous linear system

solution to a linear system

coefficient matrix of a linear system

augmented matrix of an inhomogeneous linear system

matrix in echelon form

matrix in reduced echelon form

echelon form of a matrix

reduced echelon form of a matrix

square matrix

invertible matrix

inverse of an invertible matrix

elementary matrix

upper and lower triangular matrices

Procedures you need to be familiar with are:

back substitution for solving a linear system in echelon form

Gaussian elimination

matrix inversion algorithm

Quiz

1. Solve the following **linear system**:

$$\begin{array}{rcrcrcr} 3x & - & y & - & 4z & = & 16 \\ & & 2y & + & z & = & 6 \\ & & & & 3z & = & -6 \end{array}$$

2. Solve the following **linear system**:

$$\begin{array}{rcrcrcrcr} 5x_1 & - & x_2 & + & 3x_3 & - & 4x_4 & = & 15 \\ & & 4x_2 & + & 2x_3 & - & x_4 & = & 17 \\ & & & & -3x_3 & + & x_4 & = & -3 \\ & & & & & & 2x_4 & = & 6 \end{array}$$

3. Let $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 3 \\ 2 & 5 & 2 \end{pmatrix}$. Write down a sequence $(E_1, E_2, \dots, E_k)$ of **elementary matrices** such that $(E_k \dots E_1)A$ is an **echelon form** of A.

Quiz solutions

New Concepts

We define two new concepts in this section:

a factorization or decomposition of a matrix A

LU-decomposition of a square matrix A

Theory (Why It Works)

For many applications in linear algebra it is desirable to express a matrix A as the **product** of two or more matrices. Such an expressions is called a ***factorization*** or a ***decomposition***.

Definition 3.25. Let $A, B_1, B_2, \dots, B_k$ matrices. If $A = B_1B_2 \dots B_k$ then $B_1B_2 \dots B_k$is referred to as a ***factorization or decomposition of*** A.

In most applications the **factorization** only involves a **product** of two matrices but in some cases more may be required. For example, in the mathematical field known as **Lie theory** one may factorize an **invertible** $n \times n$ matrix A as a product U_1DWU_2 where U_1, U_2 are **upper triangular matrices** with ones on their diagonal, D is an invertible **diagonal matrix** and W is a "permutation" matrix, that is, a matrix which has exactly one non-zero entry in each row and column and that entry is a one. The possibilities for W when $n = 3$ are

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

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

Such a factorizations is called the **Bruhat decomposition of A**.

Important for its practical applications in solving multiple **linear systems** with the same **coefficient matrix** A is the ***LU-decomposition***:

Definition 3.26. Let A be an $n \times n$ matrix. An $LU-$***decomposition of*** A is a **factorization** of the form $A = LU$ where L is a **lower triangular matrix** and U is an **upper triangular matrix**.

When the matrix A can be factorized as $A = LU$ with L a **lower triangular matrix** and U an **upper triangular matrix** the factorization can be used to solve a **linear system** with matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ in the following way:

Set $U\boldsymbol{x} = \boldsymbol{y} = \begin{pmatrix}y_1\\ y_1\\ \vdots\\ y_n\end{pmatrix}$. Then $A\boldsymbol{x} = (LU)\boldsymbol{x} = L(U\boldsymbol{x}) = L\boldsymbol{y}$. We can then solve $A\boldsymbol{x} = \boldsymbol{b}$ in two steps:

1. Solve $L\boldsymbol{y} = \boldsymbol{b}$.

2. Having obtained a solution $\boldsymbol{y}$ in step 1, solve the system $U\boldsymbol{x} = \boldsymbol{y}$.

Because the **coefficient matrix** of the the system in step 2 has an **upper triangular matrix** it can easily be solved by **back substitution**.

The **coefficient matrix** of the **linear system** with matrix equation $L\boldsymbol{y} = \boldsymbol{b}$ is **lower triangular matrix**. This can be solved in a way similar to **back substitution** but one solves in the order $y_1, y_2, \ldots, y_n$. We illustrate with an example.

Example 3.6.1. Performing the required multiplication validates that the matrix $A = \begin{pmatrix} 2 & -4 & 2 \\ 3 & -5 & 6 \\ -2 & 5 & 5 \end{pmatrix}$ has the LU-decomposition $A = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$. Use this to solve the linear system with matrix equation

$$A\boldsymbol{x} = A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 5 \\ -14 \end{pmatrix} \tag{3.10}$$

We can write Equation (3.10) as

$$\begin{pmatrix} 2 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 5 \\ -14 \end{pmatrix} \tag{3.11}$$

We define y_1, y_2, y_3 by $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ It then is the case that

$$\begin{pmatrix} 2 & 0 & 0 \\ 3 & 1 & 0 \\ -2 & 1 & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 5 \\ -14 \end{pmatrix} \tag{3.12}$$

This is equivalent to the linear system

$$\begin{array}{rcrcrcr} 2y_1 & & & & & = & 6 \\ 3y_1 & + & y_2 & & & = & 5 \\ -2y_1 & + & y_2 & + & 4y_3 & = & -14 \end{array} \tag{3.13}$$

We solve this like back substitution but in the order y_1, y_2, y_3. We get the unique solution $y_1 = 3, y_2 = -4$ and $y_3 = -1$. By the definition of $\boldsymbol{y}$ we have

$$\begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \\ -1 \end{pmatrix} \tag{3.14}$$

which gives the linear system

$$\begin{array}{rcrcrcr} x_1 & - & 2x_2 & + & x_3 & = & 3 \\ & & x_2 & + & 3x_3 & = & -4 \\ & & & & x_3 & = & -1 \end{array} \tag{3.15}$$

This has the unique solution $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}$ which is obtained by using **back substitution**.

Finding an $LU-$decomposition

We now determine conditions under which a **square matrix** A has an **LU-decomposition** and how to find one when it exists.

Thus, let A be an $n \times n$ matrix and let U be an **echelon form** of A obtained by the forward pass in the application of **Gaussian elimination**. Then the matrix U is an **upper triangular matrix**. We saw in Section (3.5) that there is a sequence $(E_1, E_2, \ldots, E_k)$ of **elementary matrices** such that

$$E_k E_{k_1} \ldots E_2 E_1 A = U \tag{3.16}$$

Elementary matrices are **invertible** and so we can multiply both sides by $(E_k E_{k-1} \ldots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \ldots E_{k-1}^{-1} E_k^{-1}$ to obtain

$$A = (E_1^{-1} E_2^{-1} \ldots E_{k-1}^{-1} E_k^{-1}) U \tag{3.17}$$

Since each E_i corresponds to an **elementary row operation** from the **forward pass** they are either of the form $D_s(c)$, $T_{st}(c)$ with $s < t$ or P_{st} with $s < t$. Suppose that it is unnecessary to perform any exchange operations, so that none of the E_i is of the form P_{st}. Since the matrices $D_s(c)$ are diagonal and the matrices $T_{st}(c)$ with $s < t$ are **lower triangular**, and the **product** of lower triangular matrices and diagonal matrices is lower triangular (see the **Challenge Exercises**) the matrix

$$L = E_1^{-1} E_2^{-1} \ldots E_{k-1}^{-1} E_k^{-1} \tag{3.18}$$

is a **lower triangular** and $A = LU$ is a **factorization** of A into the **product** of a lower triangular and upper triangular matrix. We illustrate with an example.

Example 3.6.2. We use the matrix $A = \begin{pmatrix} 2 & -4 & 2 \\ 3 & -5 & 6 \\ -2 & 5 & 5 \end{pmatrix}$ of **Example** (3.6.1).

The following shows a sequence of **elementary row operations** which transforms A into an **echelon form**. Beginning with the operation that scales the first row by $\frac{1}{2}$ $(R_1 \to \frac{1}{2}R_1)$ we obtain the matrix

$$\begin{pmatrix} 1 & -2 & 1 \\ 3 & -5 & 6 \\ -2 & 5 & 5 \end{pmatrix}. \tag{3.19}$$

Next, by adding -3 times the first row to the second row $(R_2 \to (-3)R_1 + R_2)$ we get the matrix

$$\begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ -2 & 5 & 5 \end{pmatrix}. \tag{3.20}$$

Then we add twice the first row to the third row $(R_3 \to 2R_1 + R_3)$

$$\begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 1 & 7 \end{pmatrix}. \tag{3.21}$$

With one more elimination operation, adding (-1) times the second row to the third row $(R_3 \to (-1)R_2 + R_3)$ we get an upper triangular matrix:

$$\begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 4 \end{pmatrix}. \tag{3.22}$$

Finally, scale the third row by $\frac{1}{4}$: $(R_3 \to \frac{1}{4}R_3)$ to get the following **upper triangular matrix** U with 1's on the diagonal:

$$\begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.23}$$

The operation in (3.19) can be achieved by multiplying A by the **elementary matrix** $D_1(\frac{1}{2})$. In (3.20) the relevant matrix is $T_{12}(-3)$. The operation in (3.21) is obtained by $T_{13}(2)$ and that in (3.22) by $T_{23}(-1)$. Finally, the last operation is achieved by multiplying by $D_3(\frac{1}{4})$. Taken together this implies that

$$[D_3(\frac{1}{4})T_{23}(-1)T_{13}(2)T_{12}(-3)D_1(\frac{1}{2})]A = U = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.24}$$

As a consequence of **Equation** (3.24) we have

$$A = [D_3(\frac{1}{4})T_{23}(-1)T_{13}(2)T_{12}(-3)D_1(\frac{1}{2})]^{-1}U =$$

$$[D_3(\frac{1}{4})T_{23}(-1)T_{13}(2)T_{12}(-3)D_1(\frac{1}{2})]^{-1}\begin{pmatrix}1 & -2 & 1\\ 0 & 1 & 3\\ 0 & 0 & 1\end{pmatrix}. \tag{3.25}$$

Recall that the **inverse** of a **product** of invertible matrices is the product of their inverses in reverse order (see **Theorem** (3.4.7) and **Remark** (3.12)).

Since $T_{ij}(c)^{-1} = T_{ij}(-c)$ and $D_i(c)^{-1} = D_i(\frac{1}{c})$ it therefore follows that

$$[(D_3(\frac{1}{4})T_{23}(-1)T_{13}(2)T_{12}(-3)D_1(\frac{1}{2})]^{-1} = D_1(2)T_{12}(3)T_{13}(-2)T_{23}(1)D_3(4) =$$

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

$$\begin{pmatrix}2 & 0 & 0\\ 3 & 1 & 0\\ -2 & 1 & 4\end{pmatrix} \tag{3.26}$$

Thus, we have obtained the decomposition given in **Example** (3.6.1).

Remark 3.18. A glance at the matrix L of **Example** (3.6.2) indicates that it is not really necessary to write out all the **elementary matrices** and multiply them together but rather the matrix L can be determined by judicious bookkeeping. First of all, the scaling operations used were $R_1 \to \frac{1}{2}R_1$ and $R_3 \to \frac{1}{4}R_3$. The diagonal entries of L occur in the first and third positions and are two and four, respectively, the reciprocals of the scaling factors. The other operations were elimination operations of the form $R_j \to c_{ij}R_i + R_j$ with $i < j$. Notice that the $(i,j)-$entry in L is $-c_{ij}$ in the respective cases. This holds in general. Consequently, L can be determined from the application of **Gaussian elimination** in the following way:

Keep track of all the **elementary row operations** . If A has an **LU-decomposition** these are all scaling and elimination operations. If the i^{th} row is rescaled by a factor c then set $l_{ii} = \frac{1}{c}$. If the i^{th} row is not rescaled during the entire process then set $l_{ii} = 1$.

If an elimination operation of the form $R_j \to dR_i + R_j$ with $i < j$ then set $l_{ij} = -d$. If at no time is a multiple of the ith row added to the j^{th} row with $i < j$ then set $l_{ij} = 1$. Finally, if $i > j$ set $l_{ij} = 0$. Then L is the matrix with entries l_{ij}.

We summarize what we have shown:

Theorem 3.6.1. *Let A be an $n \times n$ matrix. If A can be reduced to a* **row echelon form** *U without using exchange operations then there exists a* **lower triangular matrix** *L such that $A = LU$.*

What You Can Now Do

1. Given a **factorization** $A = LU$ of a **square matrix** A where L is a **lower triangular matrix** and U is an **upper triangular matrix**, solve the **linear system** with matrix equation $A\boldsymbol{x} = \boldsymbol{b}$.

2. Determine if a **square matrix** A has an **LU-decomposition**, and, if so, find such a decomposition.

Method (How To Do It)

Method 3.6.1. Given a **factorization** $A = LU$ of a **square matrix** A where L is a **lower triangular matrix** and U is an **upper triangular matrix**, solve the **linear system** with matrix equation $A\boldsymbol{x} = \boldsymbol{b}$.

Set $U\boldsymbol{x} = \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$. Solve the **linear system** with matrix equation $L\boldsymbol{y} = \boldsymbol{b}$ in a way similar to **back substitution**, solving for the variables in reverse order: find y_1 from the first equation, substitute this into the remaining equations, then find y_2 and continue until the solution is obtained.

Having found $\boldsymbol{y}$ solve the **linear system** with matrix equation $U\boldsymbol{x} = \boldsymbol{y}$ by **back substitution**.

Example 3.6.3. Use the **LU-decomposition**

$$\begin{pmatrix} 3 & 9 & 6 \\ -1 & -1 & -6 \\ 1 & 0 & 10 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ -1 & 2 & 0 \\ 1 & -3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

to solve the **linear system** with matrix equation

$$\begin{pmatrix} 3 & 9 & 6 \\ -1 & -1 & -6 \\ 1 & 0 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \\ -11 \end{pmatrix}$$

Setting $\boldsymbol{y} = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ we first solve the matrtix equation

$$\begin{pmatrix} 3 & 0 & 0 \\ -1 & 2 & 0 \\ 1 & -3 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \\ -11 \end{pmatrix}$$

This is equivalent to the **linear system**

$$\begin{array}{rcrcrcr} 3y_1 & & & & & = & 0 \\ -y_1 & + & 2y_2 & & & = & 6 \\ y_1 & - & 3y_2 & + & 2y_3 & = & -11 \end{array}$$

This has the **solution** $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ -1 \end{pmatrix}$. Now we must solve the matrix equation

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

which is equivalent to the **linear system**

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

Using **back substitution** we get the $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$.

Example 3.6.4. Use the **LU-decomposition**

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

to solve the **linear system** with matrix equation

$$\begin{pmatrix} 2 & 6 & 4 \\ 1 & 4 & 5 \\ -2 & -3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \\ 3 \end{pmatrix}$$

Set $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$. We must first solve the matrix equation

$$\begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -2 & 3 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \\ 3 \end{pmatrix}$$

which is equivalent to the **linear system**

$$\begin{array}{rcrcrcl} 2y_1 & & & & & = & 8 \\ y_1 & + & y_2 & & & = & 9 \\ -2y_1 & + & 3y_2 & - & 2y_3 & = & 3 \end{array}$$

This has the solution $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 2 \end{pmatrix}$.

We now must solve the matrix equation

$$\begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 2 \end{pmatrix}$$

which is equivalent to the **linear system**

$$\begin{array}{rcrcrcl} x_1 & + & 3x_2 & + & 2x_3 & = & 4 \\ & & x_2 & + & 3x_3 & = & 5 \\ & & & & x_3 & = & 2 \end{array}$$

By **back substitution** we get the solution $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$.

Method 3.6.2. Determine if a **square matrix** A has an **LU-decomposition**, and, if so, find such a decomposition.

Apply **Gaussian elimination** to obtain an echelon form of A. If no exchange operations are necessary in the **forward pass** (which takes A to an **upper triangular matrix** with all **leading entries** equal to 1), then A has an **LU-decomposition**, otherwise it does not.

If no exchange operations are used keep track of the operations and express them as elementary matrices $E_1, E_2, \ldots, E_k$ and let U be the matrix obtained at the end of the **forward pass**. Set $L = E_1^{-1}E_2^{-1}\ldots E_k^{-1}$. Then $A = LU$ is an **LU-decomposition** of A.

To compute L it is not actually necessary to perform the **matrix multiplications**. Rather the matrix L can be determined directly from the **Gaussian elimination** in the following way:

Keep track of all the **elementary row operations** which are scaling elimination operations.

If the i^{th} row is rescaled by a factor c then set $l_{ii} = \frac{1}{c}$. If the i^{th} row is not rescaled during the entire process then set $l_{ii} = 1$.

If an elimination operation of the form $R_j \to dR_i + R_j$ with $i < j$ then set $l_{ij} = -d$. If as no time is a multiple of the ith row added to the j^{th} row with $i < j$ then set $l_{ij} = 1$. Finally, if $i > j$ set $l_{ij} = 0$. Then L is the matrix with entries l_{ij}.

Example 3.6.5. Determine if $A = \begin{pmatrix} 1 & 2 & -2 \\ 2 & 4 & -6 \\ 3 & 5 & -8 \end{pmatrix}$ has an **LU-decomposition**.

We apply **Gaussian elimination**:

$$\begin{pmatrix} 1 & 2 & -2 \\ 2 & 4 & -6 \\ 3 & 5 & -8 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & 0 & -2 \\ 3 & 5 & -9 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & 0 & -2 \\ 0 & -1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & -1 & -3 \\ 0 & 0 & -2 \end{pmatrix} \to$$

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

The operations used, in order, are:
1) $R_2 \to (-2)R_1 + R_2$

2) $R_3 \to (-3)R_1 + R_3$

3) $R_2 \leftrightarrow R_3$

4) $R_2 \to (-1)R_2$

5) $R_3 \to (-\frac{1}{2})R_3$

The third of these is an exchange operation so this matrix does not have an **LU-decomposition**.

Example 3.6.6. Determine if the matrix $A = \begin{pmatrix} 1 & 2 & 0 \\ -4 & -6 & -4 \\ 3 & 4 & 7 \end{pmatrix}$ has an **LU-decomposition**.

The **Gaussian elimination** of this matrix through the completion of the **forward pass** is as follows:

$$\begin{pmatrix} 1 & 2 & 0 \\ -4 & -6 & -4 \\ 3 & 4 & 7 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & -4 \\ 3 & 4 & 7 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & -4 \\ 0 & -2 & 7 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & 7 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

The operations used, in order are:

1) $R_2 \to 4R_1 + R_2$

2) $R_3 \to (-3)R_1 + R_3$

3) $R_2 \to \frac{1}{2}R_2$

4) $R_3 \to 2R_2 + R_3$

5) $R_3 \to (\frac{1}{3})R_3$

These are all scaling and elimination operations and so this matrix has an **LU-decomposition**.

We can find L by multiplying the appropriate **inverses** of the **elementary matrices** in reverse order. Alternatively, we can find L directly as described in (3.6.2): the diagonal entries of L are $l_{11} = 1, l_{22} = 2, l_{33} = 3$. The entries below the main diagonal are $l_{12} = -4, l_{13} = 3, l_{23} = -2$. Therefore

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

On the other hand, the elementary matrices corresponding to the operations in 1) - 5) are: $T_{12}(4), T_{13}(-3), D_2(\frac{1}{2}), T_{23}(2), D_3(\frac{1}{3})$. Set

$$L = T_{12}(4)^{-1}T_{13}(-3)^{-1}D_2(\frac{1}{2})^{-1}T_{23}(2)^{-1}D_3(\frac{1}{3})^{-1} =$$

$$T_{12}(-4), T_{13}(3)D_2(2)T_{23}(-2)D_3(3) =$$

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

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

Thus, the **LU-decomposition** is

$$\begin{pmatrix} 1 & 2 & 0 \\ -4 & -6 & -4 \\ 3 & 4 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -4 & 2 & 0 \\ 3 & -2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

Example 3.6.7. Determine if the matrix $A = \begin{pmatrix} 1 & 2 & -2 \\ -1 & -1 & 4 \\ 1 & 1 & -3 \end{pmatrix}$ has an **LU-decomposition**. If so, find one.

We apply **Gaussian elimination** this matrix to obtain an **echelon form** which will be an **upper triangular matrix** with **leading entries** equal to 1:

$$\begin{pmatrix} 1 & 2 & -2 \\ -1 & -1 & 4 \\ 1 & 1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & 2 \\ 1 & 1 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & 2 \\ 0 & -1 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

The **elementary row operations** used are:

1) $R_2 \to R_1 + R_2$

2) $R_3 \to (-1)R_1 + R_3$

3) $R_3 \to R_2 + R_3$

The operations are all elimination operations and therefore this matrix has an **LU-decomposition**. We again find L by the direct method and by **multiplying** the appropriate **elementary matrices**.

No rescaling is necessary and therefore $l_{11} = l_{22} = l_{33} = 1$. From 1) -3) we see that $l_{12} = -1, l_{13} = 1, l_{23} = -1$. Therefore,

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

Alternatively, the **elementary matrices** corresponding to the operations 1) - 3) are:

$T_{12}(1), T_{13}(-1), T_{23}(1)$. Set

$$L = T_{12}(1)^{-1}T_{13}(-1)^{-1}T_{23}(1)^{-1} = T_{12}(-1)T_{13}(1), T_{23}(-1) =$$

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

Finally, we obtain the **LU-decomposition**

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

Exercises

In exercises 1 - 5 use the given **LU-decomposition**. to solve the **linear system** . See **Method** (3.6.1).

1. $A = \begin{pmatrix} 1 & -3 & 1 \\ 2 & -5 & 4 \\ -1 & 6 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & -3 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -4 \\ 31 \\ -2 \end{pmatrix}$

2. $A = \begin{pmatrix} 1 & 2 & 2 \\ -1 & 0 & -8 \\ 3 & 7 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ 3 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}, A\boldsymbol{x} = \begin{pmatrix} 2 \\ -4 \\ 3 \end{pmatrix}$

3. $A = \begin{pmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & -3 & 0 \\ -2 & -6 & 9 \end{pmatrix} \begin{pmatrix} 1 & -2 & -2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$, $A\boldsymbol{x} = \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}$

4. $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 6 & -2 \\ 3 & 9 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ 3 & 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$, $A\boldsymbol{x} = \begin{pmatrix} 5 \\ -8 \\ 4 \end{pmatrix}$

5. $A = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 5 & -1 \\ 3 & 9 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 3 & 0 \\ 3 & 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$, $A\boldsymbol{x} = \begin{pmatrix} -5 \\ 4 \\ -14 \end{pmatrix}$

In exercises 6 - 12 determine if the given matrix has an **LU-decomposition**. If so, find such a decomposition.

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

7. $\begin{pmatrix} 0 & 1 & 2 \\ 2 & 4 & 8 \\ 3 & 8 & 9 \end{pmatrix}$

8. $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 6 \\ 4 & 5 & 3 \end{pmatrix}$

9. $\begin{pmatrix} 2 & 4 & -4 \\ 3 & 2 & -1 \\ 2 & 8 & -6 \end{pmatrix}$

10. $\begin{pmatrix} 3 & -6 & 6 \\ 2 & -4 & 7 \\ 5 & -8 & -6 \end{pmatrix}$

11. $\begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 5 & 3 & 1 \\ 3 & 8 & 5 & -2 \\ 4 & 11 & 4 & -2 \end{pmatrix}$

12 $\begin{pmatrix} 1 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & 7 \end{pmatrix}$

Challenge Exercises (Problems)

1. Prove that the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ does not have an $LU-$decomposition.

2. For $i \neq j$ prove that $T_{ij}(a)T_{ij}(b) = T_{ij}(b)T_{ij}(a) = T_{ij}(a+b)$.

3. a) Let $D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$ and $L = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Demonstrate that DL and LD are **lower triangular**.

b) State and prove general result.

4. Assume that $i < j < k$ and a, c are scalars with $c \neq 0$. Prove that $D_i(a)T_{jk}(c) = T_{jk}(c)D_i(a)$.

5. a) Prove that the product $D_1(a_{11})T_{12}(a_{21})T_{13}(a_{31})D_2(a_{22})T_{23}(a_{32})D_3(a_{33}) = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$

b) State a similar result for arbitrary n and explain how this relates to a direct determination of the matrix L from the application of **Gaussian elimination** to a matrix A which has an **LU-decomposition**.

Quiz Solutions

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

Not right, see **back substitution**.

2. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 2 \\ 3 \end{pmatrix}$.

Not right, see **back substitution**.

3. Let $E_1 = T_{12}(-2) = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$,

$E_2 = T_{13}(-2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$,

$E_3 = D_2(-1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, and

$E_4 = T_{23}(-1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$.

Then $E_4E_3E_2E_1A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$.

Not right, see **Gaussian elimination** and **Method** (3.5.1).

3.7. How to Use It: Applications of Matrix Multiplication

In this section we show how the algebra of matrices can be used to model situations in such areas as ecology, population, and economics.

Stochastic Processes and Markov Chain Models

A ***stochastic process*** is a process that has a finite number of ***outcomes or states*** and at any one time there is a specific probability that the process is in a particular state. We will label the states by the natural numbers $1, 2, \ldots, n$ and the probability that the system is in state i by p_i. Then the vector $\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}$ is called a ***state vector***. Note that all the entries of a state vector are nonnegative and their sum is one. A state vector is therefore an example of a probability vector:

Definition 3.27. A vector $\boldsymbol{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}$ which satisfies

1) $p_i \geq 0$ for all i; and

2) $p_1 + p_2 + \cdots + p_n = 1$ is called a ***probability vector***

We shall be interested in $n \times n$ matrices whose columns are probability vectors. These matrices have a special name:

Definition 3.28. A **square matrix** T whose columns are all **probability vectors** is called a ***stochastic matrix***.

Example 3.7.1. The matrix $\begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{4} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{12} & \frac{1}{3} \end{pmatrix}$ is a **stochastic matrix** and the vector $\begin{pmatrix} \frac{1}{6} \\ \frac{1}{2} \\ \frac{1}{3} \end{pmatrix}$ is a **probability vector**. Let us compute the product:

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{4} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{12} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{6} \\ \frac{1}{2} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{23}{72} \\ \frac{35}{72} \\ \frac{7}{36} \end{pmatrix}.$$

Since $\frac{23}{72} + \frac{35}{72} + \frac{7}{36} = 1$ the resulting **3-vector** is a **probability vector**. This always happens as stated in our first result:

Theorem 3.7.1. *Let T be an $n \times n$ **stochastic matrix** and $\boldsymbol{p}$ a **probability vector** from $\mathbb{R}^n$. Then $T\boldsymbol{p}$ is a probability vector.*

We leave this as a **challenge exercise**. An immediate consequence of this is the following theorem about powers of a **stochastic matrix**:

Theorem 3.7.2. *Let T be a **stochastic matrix**. Then for each positive integer k, T^k is a stochastic matrix.*

We also assign this as a **challenge exercise**.

We can now define what we mean by a Markov chain:

Definition 3.29. A ***Markov chain*** consists of a sequence $(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots)$ of **state vectors** and a **stochastic matrix** T, called the ***transition matrix***, such that for every k, $\boldsymbol{x}_{k+1} = T\boldsymbol{x}_k$.

Think of a **Markov chain** as modeling some process that changes over time with the state of the process recorded at discrete intervals of equal duration. This simple idea can be used to analyze situations in economics, business, psychology, engineering, politics, marketing, ecology and many more areas and domains. We illustrate with a couple of examples.

Example 3.7.2. A Markov Model of Gypsy Moths

The gypsy moth is an extremely destructive organism which has wreaked havoc in North American forests. Since introduced in the northeastern United States around 1870 it has expanded over nearly 500,000 square miles. The resulting defoliation has had a significant economic and ecological effect.

It is important to predict future occurrences of gypsy moth outbreaks and to do so, ecologists have attempted to model the spatial dynamics of its spread and effect. In such a model the researchers divide a forested region into cells of a given size and observe these over many time periods of equal duration. For any particular time period they are classified as state 1, nondefoliated, or state 2, defoliated.

Let us suppose that some forest region is divided into 1,000 cells. The state of this region can be summarized by the number of cells which are nondefoliated and the number that are defoliated together in a single vector. For example, $\boldsymbol{v}_0 = \begin{pmatrix} 800 \\ 200 \end{pmatrix}$ is an example of a vector which might describe our hypothetical forest as initially observed.

Based on observation from previous investigations the researchers hypothesize probabilities that a cell with a given state in one period will be in either of the two states in the following period. Specifically, they hypothesized that if a cell was nondefoliated in one period there was roughly a 0.95 probability that it would be nondefoliated in the subsequent period and 0.05 probability that it would be defoliated. Similarly, if the cell was defoliated in one period then the probability that it would be nondefoliated in the next period was given as 0.65 and 0.35 that it would continue to be defoliated. The fact that the probability that a defoliated area is more likely to become nondefoliated than continue to be defoliated reflects the fact that defoliation recedes faster than it persists.

We can formulate this as a **Markov chain**. We obtain an initial state the vector $\boldsymbol{v}_0$ by dividing by 1000, the number of total cells in the sample. The resulting vector gives the probabilities that a cell within this forest region is nondefoliated or defoliated as its components: $\boldsymbol{x}_0 = \frac{1}{1000}\boldsymbol{v}_0 = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}$.

The **transition matrix** for this process is

$$T = \begin{pmatrix} 0.95 & 0.65 \\ 0.05 & 0.35 \end{pmatrix}$$

We obtain a **Markov chain** by setting $\boldsymbol{x}_1 = T\boldsymbol{x}_0, \boldsymbol{x}_2 = T\boldsymbol{x}_1$ and in general, $\boldsymbol{x}_{k+1} = T\boldsymbol{x}_k$. If we wish to compute the expected number of cells which are nondefoliated, respectively, defoliated, we can multiply $\boldsymbol{x}_k$ by 1000 to obtain $\boldsymbol{v}_k$. We compute the first couple of **state vectors**

$$\boldsymbol{x}_1 = T\boldsymbol{x}_0 = \begin{pmatrix} 0.95 & 0.65 \\ 0.05 & 0.35 \end{pmatrix} \begin{pmatrix} 0.80 \\ 0.20 \end{pmatrix} = \begin{pmatrix} 0.89 \\ 0.11 \end{pmatrix}$$

$$\boldsymbol{x}_2 = T\boldsymbol{x}_1 = \begin{pmatrix} 0.95 & 0.65 \\ 0.05 & 0.35 \end{pmatrix} \begin{pmatrix} 0.89 \\ 0.11 \end{pmatrix} = \begin{pmatrix} 0.917 \\ 0.083 \end{pmatrix}.$$

To get a good idea of how the chain evolves, let's investigate the values of the state vector over ten iterations:

$$\boldsymbol{x}_3 = T\boldsymbol{x}_2 = \begin{pmatrix} 0.95 & 0.65 \\ 0.05 & 0.35 \end{pmatrix} \begin{pmatrix} 0.917 \\ 0.083 \end{pmatrix} = \begin{pmatrix} 0.9251 \\ 0.0749 \end{pmatrix}$$

$$\boldsymbol{x}_4 = T\boldsymbol{x}_3 = \begin{pmatrix} 0.95 & 0.65 \\ 0.05 & 0.35 \end{pmatrix} \begin{pmatrix} 0.9251 \\ 0.0749 \end{pmatrix} = \begin{pmatrix} 0.92753 \\ 0.07247 \end{pmatrix}$$

$$\boldsymbol{x}_5 = T\boldsymbol{x}_4 = \begin{pmatrix} 0.95 & 0.65 \\ 0.05 & 0.35 \end{pmatrix} \begin{pmatrix} 0.92753 \\ 0.07247 \end{pmatrix} = \begin{pmatrix} 0.928259 \\ 0.071741 \end{pmatrix}$$

$$\boldsymbol{x}_{10} = T\boldsymbol{x}_9 = \begin{pmatrix} 0.95 & 0.65 \\ 0.05 & 0.35 \end{pmatrix} \begin{pmatrix} 0.928569 \\ 0.0714311 \end{pmatrix} = \begin{pmatrix} 0.928569 \\ 0.0714311 \end{pmatrix}$$

The vector $\boldsymbol{x}_{10}$ appears to be fixed by T (it is only approximately so, since we have truncated the decimal). However, there is a vector which is fixed, namely, $\begin{pmatrix} \frac{13}{14} \\ \frac{1}{14} \end{pmatrix}$ (check this). Such a vector is an example of a ***steady state vector***.

Definition 3.30. Let T be a **stochastic matrix** . If the **probability vector** $\boldsymbol{p}$ satisfies $T\boldsymbol{p} = \boldsymbol{p}$ then it is called a ***steady-state vector for*** T.

In **Example** (3.7.2) we can show that the vector $\begin{pmatrix} \frac{13}{14} \\ \frac{1}{14} \end{pmatrix}$ is the unique steady state vector. For suppose $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is a steady state vector. Then $x_1, x_2 \geq 0, x_1 + x_2 = 1$ and

$$\begin{pmatrix} \frac{19}{20} & \frac{13}{20} \\ \frac{1}{20} & \frac{7}{20} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{3.27}$$

Performing the matrix multiplication on the left hand side of (3.27) we get

$$\begin{pmatrix} \frac{19}{20}x_1 + \frac{13}{20}x_2 \\ \frac{1}{20}x_1 + \frac{7}{20}x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{3.28}$$

Now subtract $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ from both sides in (3.28):

$$\begin{pmatrix} -\frac{1}{20}x_1 + \frac{13}{20}x_2 \\ \frac{1}{20}x_1 - \frac{13}{20}x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.29}$$

Thus, $\boldsymbol{x}$ is a **solution** to the **homogeneous linear system** with **coefficient matrix** $\begin{pmatrix} -\frac{1}{20} & \frac{13}{20} \\ \frac{1}{20} & -\frac{13}{20} \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & -13 \\ 0 & 0 \end{pmatrix}$ and the **solution set** to the system is $Span\left(\begin{pmatrix} 13 \\ 1 \end{pmatrix}\right)$. If $x_1 = 13t, x_2 = t$ and $x_1 + x_2 = 1$ then $t = \frac{1}{14}$ as claimed.

We will see in the chapter on eigenvectors and eigenvalues of matrices that every **stochastic matrix** has a **steady state vector**. Moreover, if the transition matrix is regular (defined immediately below) then it can be shown there there is a unique steady state vector.

Definition 3.31. A **stochastic matrix** T is said to be ***positive*** if every entry of T is positive. T is said to ***regular*** for some natural number k if the matrix T^k is positive.

The importance of **regular stochastic matrices** is indicated by the following theorem:

Theorem 3.7.3. *Let T be a* **regular stochastic matrix**. *Then the powers T^k of T approach a matrix S of the form $(\boldsymbol{p}\,\boldsymbol{p}\ \dots\ \boldsymbol{p})$ where $\boldsymbol{p}$ is a* **steady state vector** *for T. Moreover, any vector $\boldsymbol{x}$ which satisfies $T\boldsymbol{x} = \boldsymbol{x}$ is a scalar multiple of $\boldsymbol{p}$.*

We used the expression "the powers T^k of T approach S." By this we mean that for any small number ϵ there is a natural number N such that if $k \geq N$ then each entry of T^n differs from the corresponding entry of S by less than ϵ.

We do a second example.

Example 3.7.3. Hypothetical Weather Example

A meteorologist studying the weather in a particular town classified the days as Sunny = State 1, Cloudy = State 2, and Rainy = State 3. Through extensive observation she found probabilities w_{ij} that the weather is in state i on a given day after being in state j on the previous day. These probabilities are given in the following **stochastic matrix**:

$$W = \begin{pmatrix} 0.7 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.4 \\ 0.1 & 0.3 & 0.4 \end{pmatrix}$$

Find a **steady state vector** for this **Markov chain**.

If $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ is a **steady state vector** then $W\boldsymbol{x} = \boldsymbol{x}$ or $W\boldsymbol{x} - \boldsymbol{x} = (W - I_3)\boldsymbol{x} = \mathbf{0}_3$ so that $\boldsymbol{x}$ is in the **null space** of the matrix $W - I_3$.

$$W - I_3 = \begin{pmatrix} -0.3 & 0.2 & 0.2 \\ 0.2 & -0.5 & 0.4 \\ 0.1 & 0.3 & -0.6 \end{pmatrix}.$$

The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -\frac{18}{11} \\ 0 & 1 & -\frac{16}{11} \\ 0 & 0 & 0 \end{pmatrix}$.

The **null space** is **spanned by** the vector $\begin{pmatrix} \frac{18}{11} \\ \frac{16}{11} \\ 1 \end{pmatrix}$.

In order to obtain a **probability vector** we must choose a t such that

$$\frac{18}{11}t + \frac{16}{11}t + t = 1$$

from which we deduce that $t = \frac{11}{45}$. Thus, our **steady state vector** is $\begin{pmatrix} \frac{18}{45} \\ \frac{16}{45} \\ \frac{11}{45} \end{pmatrix} \sim \begin{pmatrix} 0.4 \\ 0.36 \\ 0.24 \end{pmatrix}$.

Age Structured Population Models

Often ecologists desire more fine grained data about a species then its sheer numbers and the changes in the total population over time. Rather, to get a reasonably accurate picture of the population dynamics it is necessary to take into account the distribution of the population amongst different age groups distinguished by their stages of development and position in the life cycle. A model that makes use of matrices was introduced in a pair of scholarly papers by P.H. Leslie in the 1940s.

In this model one assumes that the species population is divided into n age groups, labeled by the natural numbers $i = 1, 2, \ldots, n$ where age group i consists of the individuals whose age a satisfies $i-1 \leq a < i$. Typically, such models focus on the female population since they are the only ones which breed. The population of each group is to be followed in discrete units of time, typically the interval between breeding, usually quoted as years. The numbers of each age group are then recorded just prior to breeding. Moreover, the models incorporate the following assumptions:

1) There is a **maximum age span** for the species and consequently, every individual dies off by age n.

2) For each age group $i = 1, 2, \ldots, n$ there is a **constant birth rate** (fecundity), b_i, stated as average number of female off-spring per individual in the age group.

3) For each age group i there is a **constant mortality rate**, or the equivalent, a **constant survival rate**, s_i, which is the likelihood that an individual in the i^{th} group in period k survives to become a member of age group $i + 1$ in period $k + 1$.

This is summarized in the following table

Table 3.1: Birth and Survival Rates

Age	1	2	3	...	n
Age Class	0 -1	1- 2	2 - 3	...	(n-1) - n
Fecundity	b_1	b_2	b_3	...	b_n
Survival rate	s_1	s_2	s_3	...	-

We can define a sequence of $n-$vectors $\boldsymbol{p}(k)$ where the i^{th} component, $\boldsymbol{p}_i(k)$, of $\boldsymbol{p}(k)$ is the population of group i just prior to the breeding at the beginning of the k^{th} interval.

From these assumptions we can determine the population vector $\boldsymbol{p}(k+1)$ from $\boldsymbol{p}(k)$:

First of all, after breeding, the expected number of births in year k, and hence the population of group 1 in year $k+1$, is

$$p_1(k+1) = b_1p_1(k) + b_2p_2(k) + \cdots + b_np_n(k) \tag{3.30}$$

Since, on average, s_1 of those in age group one survive to age group 2 we have

$$p_2(k+1) = s_1p_1(k) = s_1(b_1p_1(k-1) + b_2p_2(k-1) + \cdots + b_np_n(k-1)) =$$

$$s_1b_1p_1(k-1) + s_1b_2p_2(k-1) + \cdots + s_1b_np_n(k-1) \tag{3.31}$$

Given that, on average, s_i of the individuals in age group i survive to age group $i+1$ we have

$$p_2(k+1) = s_1p_1(k), p_3(k+1) = s_2p_2(k), \ldots, p_n(k+1) = s_{n-1}p_{n-1}(k) \tag{3.32}$$

This can be written as a matrix equation:

Let

$$S = \begin{pmatrix} b_1 & b_2 & b_3 & \cdots & b_{n-1} & b_n \\ s_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & s_2 & 0 & \ldots & 0 & 0 \\ 0 & 0 & s_3 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & s_{n-1} & 0 \end{pmatrix} \tag{3.33}$$

Table 3.2: Birth and Survival Rates for Hypothetical Rodent Species

Age	0	1	2	3
Age class	0 - 1	1 - 2	2 -3	3 - 4
Birth rate	0	1	2	2
Survival rate	0.3	0.5	0.4	-

The matrix S is called the ***population projection matrix***. Notice that it has a particular shape. A **square matrix** S which has nonnegative entries and has the shape of the matrix in (3.33), that is, the only non-zero entries occur in positions $(1, j), j \geq 2$ and $(i, i-1), i = 2, 3, \ldots, n$, is called a **Leslie matrix**. We can express the relations in (3.31) and (3.32) in the single matrix equation:

$$\boldsymbol{p}(k+1) = S\boldsymbol{p}(k) \tag{3.34}$$

We illustrate with a couple of examples.

Example 3.7.4. A hypothetical species of rabbits has a maximum lifespan of four years. The females becomes sexually active at one year. At age one she is likely to produce 1.0 female offspring, at two years 3.0 female offspring and at three years 2.0 female offspring. The survival rates, determined by predators and food supply, are: $s_1 = 0.3, s_2 = 0.5, s_3 = 0.4$. This is summarized in table two above.

The **population projection matrix** for this hypothetical population is

$$S = \begin{pmatrix} 0 & 1 & 2 & 2 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0.4 & 0 \end{pmatrix}$$

Assume that initially there are 400 rabbits of age 0 -1, 200 age 1 - 2, 100 age 2 - 3 and and 100 of age 3 - 4, so the **initial population vector** is $\boldsymbol{p}(0) = \begin{pmatrix} 400 \\ 200 \\ 100 \\ 100 \end{pmatrix}$. Then in year one the distribution is given by

$$\boldsymbol{p}(1)=S\boldsymbol{p}(0)=\begin{pmatrix}0&1&2&2\\0.3&0&0&0\\0&0.4&0&0\\0&0&0.4&0\end{pmatrix}\begin{pmatrix}400\\200\\100\\100\end{pmatrix}=\begin{pmatrix}600\\120\\80\\40\end{pmatrix}$$

Iterating we obtain the following population vectors:

$$\boldsymbol{p}(2)=\begin{pmatrix}360\\180\\48\\32\end{pmatrix},\boldsymbol{p}(3)=\begin{pmatrix}360\\108\\72\\20\end{pmatrix},\boldsymbol{p}(4)=\begin{pmatrix}292\\102\\43\\29\end{pmatrix},\boldsymbol{p}(5)=\begin{pmatrix}246\\88\\41\\17\end{pmatrix}$$

$$\boldsymbol{p}(10)=\begin{pmatrix}106\\38\\18\\8\end{pmatrix},\boldsymbol{p}(15)=\begin{pmatrix}46\\17\\8\\4\end{pmatrix},\boldsymbol{p}(20)=\begin{pmatrix}21\\7\\4\\2\end{pmatrix},\boldsymbol{p}(25)=\begin{pmatrix}10\\3\\2\\1\end{pmatrix}$$

$$\boldsymbol{p}(30)=\begin{pmatrix}4\\1\\1\\0\end{pmatrix}.$$

This population continually decreases until it becomes extinct (after more than 40 iterations).

Example 3.7.5. A hypothetical rodent species has a maximum lifespan of three years. The female becomes sexually active at one year. At age one she is likely to produce 7 female offspring, at two years 5 female offspring. The survival rates, determined by predators and food supply, are: $s_1 = 0.50$ and $s_2 = 0.4$ This is summarized in the Table (3.3) above:

The **population projection matrix** for this hypothetical population is

$$S=\begin{pmatrix}0&7&5\\0.5&0&0\\0&0.4&0\end{pmatrix}$$

Table 3.3: Birth and Survival Rates for Hypothetical Rodent Species

Age class	1	2	3
Age range	0 - 1	1 - 2	2 - 3
Birth rate	0	7	5
Survival rate	0.5	0.4	0

Assume that initially there are 300 rodents in age class 1 (0-1 years of age), 80 in age class two and 40 in age class 3. Then the **initial population vector** is $\boldsymbol{p}(0) = \begin{pmatrix} 300 \\ 80 \\ 40 \end{pmatrix}$.
Then in year one the distribution is given by

$$\boldsymbol{p}(1) = S\boldsymbol{p}(0) = \begin{pmatrix} 0 & 7 & 5 \\ 0.5 & 0 & 0 \\ 0 & .4 & 0 \end{pmatrix} \begin{pmatrix} 300 \\ 80 \\ 40 \end{pmatrix} = \begin{pmatrix} 760 \\ 150 \\ 32 \end{pmatrix}$$

Iterating we obtain the following **population vectors**:

$$\boldsymbol{p}(2) = \begin{pmatrix} 1210 \\ 380 \\ 60 \end{pmatrix}, \boldsymbol{p}(3) = \begin{pmatrix} 2960 \\ 605 \\ 152 \end{pmatrix}$$

$$\boldsymbol{p}(4) = \begin{pmatrix} 4995 \\ 1480 \\ 242 \end{pmatrix}, \boldsymbol{p}(5) = \begin{pmatrix} 11570 \\ 2497 \\ 592 \end{pmatrix}$$

The next five population vectors are

$$\begin{pmatrix} 20439 \\ 5785 \\ 998 \end{pmatrix}, \begin{pmatrix} 45485 \\ 10220 \\ 2314 \end{pmatrix}, \begin{pmatrix} 83110 \\ 22743 \\ 4088 \end{pmatrix}, \begin{pmatrix} 179641 \\ 41555 \\ 9097 \end{pmatrix}, \begin{pmatrix} 336370 \\ 89821 \\ 16622 \end{pmatrix} \tag{3.35}$$

Clearly all the age classes are increasing and quite rapidly. The sequence of the total populations is: 420, 942, 1650, 3717, 6717, 14659, 27222, 58019, 109941, 230293, 442813. Dividing each number by its predecessor we obtain the growth rates from year to year: 2.24, 1.75, 2.25, 1.81, 2.18, 1.86, 2.13, 1.89, 2.09, 1.92. If we continued this way we would see that these numbers approach 2.

It is also worthwhile to look at the percentage that each class makes of the total population. The vectors we obtain are the ***relative population vectors***. For the given sample we obtain the vectors:

$$\hat{\boldsymbol{p}}(0) = \frac{1}{420}\boldsymbol{p}(0) = \begin{pmatrix} .71 \\ .19 \\ .10 \end{pmatrix}, \hat{\boldsymbol{p}}(1) = \frac{1}{942}\boldsymbol{p}(1) = \begin{pmatrix} .81 \\ .16 \\ .03 \end{pmatrix}, \hat{\boldsymbol{p}}(2) = \frac{1}{1650}\boldsymbol{p}(2) = \begin{pmatrix} .73 \\ .23 \\ .04 \end{pmatrix}$$

$$\hat{\boldsymbol{p}}(3) = \frac{1}{3717}\boldsymbol{p}(3) = \begin{pmatrix} .80 \\ .16 \\ .04 \end{pmatrix}, \hat{\boldsymbol{p}}(4) = \frac{1}{6717}\boldsymbol{p}(4) = \begin{pmatrix} .74 \\ .22 \\ .04 \end{pmatrix}, \hat{\boldsymbol{p}}(5) = \frac{1}{14659}\boldsymbol{p}(5) = \begin{pmatrix} .79 \\ .17 \\ .04 \end{pmatrix}.$$

$$\hat{\boldsymbol{p}}(6) = \begin{pmatrix} .75 \\ .21 \\ .04 \end{pmatrix}, \hat{\boldsymbol{p}}(7) = \begin{pmatrix} .78 \\ .18 \\ .04 \end{pmatrix}, \hat{\boldsymbol{p}}(8) = \begin{pmatrix} .76 \\ .21 \\ .03 \end{pmatrix}, \hat{\boldsymbol{p}}(9) = \begin{pmatrix} .78 \\ .18 \\ .04 \end{pmatrix}, \hat{\boldsymbol{p}}(10) = \begin{pmatrix} .76 \\ .20 \\ .04 \end{pmatrix}$$

If we continued we would find that the relative population vectors approach the vector $\begin{pmatrix} \frac{10}{13} \\ \frac{5}{26} \\ \frac{1}{26} \end{pmatrix} \sim \begin{pmatrix} .77 \\ .19 \\ .04 \end{pmatrix}$.

By a ***stable age distribution vector*** we mean a **population vector** $\boldsymbol{p}$ for which the proportion of each age class remains constant. The vector $\boldsymbol{p} = \begin{pmatrix} 20 \\ 5 \\ 1 \end{pmatrix}$ is a stable age distribution vector and for this vector $S\boldsymbol{p} = 2\boldsymbol{p}$. We will understand this example better after we study eigenvalues and eigenvectors.

Example (3.7.5) is unrealistic since no population continues to increase at a fixed rate (in particular doubling) indefinitely - some kind of pressure in the form of increased predators, competition for food, or disease due to overcrowding - will reduce the growth rate and perhaps even cause the population to decline. On the other hand, it is more realistic that there is a population distribution $\boldsymbol{p}$ for a **population projection matrix** S which is invariant under S, that is, $S\boldsymbol{p} = \boldsymbol{p}$. When there exists such a vector, if at any time $\boldsymbol{p}(t)$ is a multiple of $\boldsymbol{p}$ then $\boldsymbol{p}(t+1) = \boldsymbol{p}(t+2) = \dots$ for ever after and the population remains constant. Such a vector $\boldsymbol{p}$ is called a ***stationary vector*** for S.

To see that this assertion holds, assume that $S\boldsymbol{p} = \boldsymbol{p}$ and that $\boldsymbol{p}(t) = \alpha\boldsymbol{p}$. Then

$$\boldsymbol{p}(t+1) = S(\alpha\boldsymbol{p}) = \alpha(S\boldsymbol{p}) = \alpha\boldsymbol{p} = \boldsymbol{p}(t).$$

By **mathematical induction** we will have $\boldsymbol{p}(t+k) = \boldsymbol{p}(t)$ for every natural number k. This is assigned as a **challenge exercise**.

Example 3.7.6. Consider a population with a Leslie matrix $S =$

$$\begin{pmatrix} 0 & 1 & 2 \\ \frac{3}{5} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{pmatrix}$$

Determine if this **population projection matrix** has a **stationary vector**.

Suppose $\boldsymbol{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ is a **stationary vector**. Then $\boldsymbol{p} = S\boldsymbol{p}, \boldsymbol{p} - S\boldsymbol{p} = \mathbf{0}_3$ and therefore $(I_3 - S)\boldsymbol{p} = \mathbf{0}_3$. This means that $\boldsymbol{p}$ is in **null space** of $(I_3 - S)$.

$$I_3 - S = \begin{pmatrix} 1 & -1 & -2 \\ -\frac{3}{5} & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{pmatrix}$$

The **reduced echelon form** of this matrix is

$$\begin{pmatrix} 1 & 0 & -5 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix}$$

from which we conclude that $\begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix}$ is a **stationary vector**.

Therefore, if at some point the population is in the ratio of 5: 3 : 1 then the total population will remain constant and distributed in the proportions $\frac{5}{9}, \frac{1}{3}, \frac{1}{9}$ among the age classes forever.

Leontief Input-Output Models

Wassily Leontif, building on ideas pioneered by the eighteenth century French economist **Francois Quesnay**, developed a method for analyzing economies (local, regional, national, international) known as the "input-output" method. He used this to model the United States economy with a system of 500 equations and 500 variables. For his work, Leontief was awarded the **Nobel Prize in economics** in 1973. The methodology has since been extended to incorporate exchanges between the economy and the larger physical environment.

An input-output model for an economy (local, state, regional, national, international) begins with empirical research. The economy is divided into segments or productive sectors. These could be individual industries in moderately large disaggregated models. The sectors could be even smaller, down to producers of particular products in totally disaggregated models. Or, they could be aggregated into large inhomogeneous entities such as the entire manufacturing sector. Then one determines the flow, usually denominated in dollars, from each sector (as a producer) to each sector (as a consumer) over some time period, ordinarily a year. The model is said to be **closed** if all the product is consumed by the producers and **open** if there are non-producers (government, households, "foreign consumers") which consume some of the final product.

In the general model we assume there are n sectors, $S_1, S_2, \dots, S_n$ and that we know the percentage c_{ij} of the product of sector S_j consumed by Sector S_i. In this way we obtain a matrix

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \dots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} \tag{3.36}$$

called the **consumption matrix**. In a ***closed economy***, since there is no external demand, all the product is consumed by these sectors and in this case the column sums of this matrix all add to 1, that is, for each $j = 1, 2, \dots, n$.

$$c_{1j} + c_{2j} + \dots + c_{nj} = 1 \tag{3.37}$$

Suppose for the moment that we are analyzing a closed economy and let the total output (value) of sector S_i be P_i. Leontief demonstrated that in a closed economy there is a set of prices (values P_i) that can be assigned to the outputs of the sectors that are in ***equilibrium*** - the income of each sector is equal to its costs. This means for each $i = 1, 2, \dots, n$ the following equations holds:

$$P_i = c_{i1}P_1 + c_{i2}P_2 + \dots + c_{in}P_n \tag{3.38}$$

Set $\boldsymbol{P} = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{pmatrix}$. This is the ***production vector*** and an example of an ***equilibrium vector***. The equilibrium equations (3.38) can be combined into a single matrix equation:

Table 3.4: A Three Sector Closed Economy

		Produced by Manufacturing	Agriculture	Services
Consumed by	Manufacturing	0.5	0.3	0.4
	Agriculture	0.2	0.1	0.2
	Services	0.3	0.6	0.4

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \dots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{pmatrix} \tag{3.39}$$

Leontief's result is in part a consequence of the following:

Theorem 3.7.4. *Let C be an $n \times n$ matrix such that the sum of the entries in each column is one. Then the $null(C - I_n) \neq \{\mathbf{0}_n\}$.*

This will be proved after we introduce eigenvalues and eigenvectors in chapter 7.

We demonstrate with a simple example.

Example 3.7.7. Suppose an economy is aggregated into manufacturing, agriculture and services. We assume that the output of each sector is distributed as given by table (3.4)

Thus, the **consumptiom matrix** for this economy is

$$C = \begin{pmatrix} 0.5 & 0.3 & 0.4 \\ 0.2 & 0.1 & 0.2 \\ 0.3 & 0.6 & 0.4 \end{pmatrix}$$

Let P_M denote the price (value) of the manufacturing sector, P_A the agricultural sector and P_S the service sector and $\boldsymbol{P} = \begin{pmatrix} P_M \\ P_A \\ P_S \end{pmatrix}$. If $\boldsymbol{P}$ is an **equilibrium vector** then

$$C\boldsymbol{P} = \boldsymbol{P}$$

We deduce from this that $C\boldsymbol{P} - \boldsymbol{P} = C\boldsymbol{P} - I_3\boldsymbol{P} = (C - I_3)\boldsymbol{P} = O_{3\times 3}$.

Thus, $\boldsymbol{P}$ is in the **null space** of the matrix $C - I_3 = \begin{pmatrix} -0.5 & 0.3 & 0.4 \\ 0.2 & -0.9 & 0.2 \\ 0.3 & 0.6 & -0.6 \end{pmatrix}$.
Notice that the sum of the entries in each columns is the **zero vector** and therefore the vector $\boldsymbol{j}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is in the **null space** of the transpose of $C - I_3$. Therefore $C - I_3$ is **non- invertible** and, consequently, $C - I_3$ is non-invertible. By **Gaussian elimination** we obtain the **reduced echelon form** of this matrix:

$$\begin{pmatrix} 1 & 0 & -\frac{14}{13} \\ 0 & 1 & -\frac{6}{13} \\ 0 & 0 & 0 \end{pmatrix}$$

Therefore an **equilibrium vector** has the form $\begin{pmatrix} \frac{14}{13}t \\ \frac{6}{13}t \\ t \end{pmatrix}$. We can multiply by 13 to clear fractions. Then $\begin{pmatrix} 14t \\ 6t \\ 13t \end{pmatrix}$ is an equilibrium vector.

Open Economy

Now consider an open economy. In this situation some of the output of the sectors exceeds the total of the inputs and is consumed externally. This is represented by a **final demand vector**, $\boldsymbol{D}$, whose components consist of the value of each sector that is consumed by nonproducers. If $\boldsymbol{P}$ is the production vector then the internal or **intermediate demand** vector is $C\boldsymbol{P}$ where C is the consumption matrix. Therefore if there is a balance (equilibrium) between what is produced and what is consumed, then the following matrix equation must hold:

$$\boldsymbol{P} = C\boldsymbol{P} + \boldsymbol{D} \tag{3.40}$$

Subtracting $C\boldsymbol{P}$ from both sides of (3.40) we obtain

$$\boldsymbol{P} - C\boldsymbol{P} = \boldsymbol{D}, (I_n - C)\boldsymbol{P} = \boldsymbol{D} \tag{3.41}$$

We require a solution to (3.41) for which every component of $\boldsymbol{D}$ is greater than equal to zero. This does not always occur (though it will for all 2×2 consumption matrices representing an open economy). However, if $I_n - C$ is **invertible** and $(I_n - C)^{-1}$ has nonzero entries then there will be a unique solution $\boldsymbol{P} = (I_n - C)^{-1}\boldsymbol{D}$. We illustrate with an example.

Example 3.7.8. Suppose a three sector economy of manufacturing, agriculture and services has the following **consumption matrix**

$$C = \begin{pmatrix} 0.5 & 0.3 & 0.3 \\ 0.1 & 0.5 & 0.2 \\ 0.2 & 0.1 & 0.3 \end{pmatrix} \tag{3.42}$$

Further assume that the final demand is for 1000 units of manufacturing, 600 units of agriculture and 1200 units of services. Find the necessary production level, $\boldsymbol{P}$, which will satisfy these requirements of the economy.

We form the matrix $I_3 - C =$

$$\begin{pmatrix} 0.5 & -0.3 & -0.3 \\ -0.1 & 0.5 & -0.2 \\ -0.2 & -0.1 & 0.7 \end{pmatrix}$$

The **augmented matrix** of the **linear system** we need to solve is

$$\begin{pmatrix} 0.5 & -0.3 & -0.3 & | & 1000 \\ -0.1 & 0.5 & -0.2 & | & 600 \\ -0.2 & -0.1 & 0.7 & | & 1200 \end{pmatrix}$$

By **Gaussian elimination** we obtain the **reduced echelon form**:

$$\begin{pmatrix} 1 & 0 & 0 & | & \frac{22000}{3} \\ 0 & 1 & 0 & | & \frac{40000}{9} \\ 0 & 0 & 1 & | & \frac{40000}{9} \end{pmatrix}$$

Therefore, the desired production vector is $\begin{pmatrix} \frac{22000}{3} \\ \frac{40000}{9} \\ \frac{40000}{9} \end{pmatrix}$.

It is sometimes possible to recognize when a **consumption matrix** for an open economy will have a unique solution, specifically, when the matrix is "productive".

Definition 3.32. An $n \times n$ **consumption matrix** C is said to be ***productive*** if $(I_n - C)^{-1}$ exists and all the entries in $(I_n - C)^{-1}$ are non-negative.

Under certain reasonable conditions on a productive consumption matrix C the matrix $(I_n - C)^{-1}$ can be approximated.

Consider the product of $I_n - C$ and $I_n + C + C^2 + \cdots + C^{k-1}$. Applying the distributive property we get

$$[I_n + C + C^2 + \cdots + C^{k-1}] - [C + C^2 + \cdots + C^{k-1} + C^k] = I_n - C^k$$

If the entries of C^k can be made very small then $I_n + C + \cdots + C^{k-1}$ will be a good approximation of $(I_n - C)^{-1}$. This is, indeed, the case when all the column sums of C are less than one:

Theorem 3.7.5. *Let C be a* **consumption matrix** *and assume that all the column sums of C are strictly less than one. Then for any positive number ϵ there is an n such that all the entries of C^n are less than ϵ.*

Proof. Let M be the maximum column sum of C so that $0 < M < 1$. Also, let a be the maximum among all the entries of C. We claim that all the entries in the matrix C^k are less than or equal to $M^{k-1}a$. We prove this by **mathematical induction**. We point out that since C is a **consumption matrix** all its entries are non-negative and as a consequence this is also true of all powers C^k of C, k a natural number.

The base or initial case is $k = 2$ and so we must show that every entry of C^2 is less than or equal to Ma. Now the $(i, j)-$entry of C^2 is the product of the i^{th} row and j^{th} column of C. Represent these by $\boldsymbol{r}_i$ and $\boldsymbol{c}_j$.

Since each entry of $\boldsymbol{c}_j$ is non-negative and each entry of $\boldsymbol{r}_i$ is at most a it follows that $\boldsymbol{r}_i\boldsymbol{c}_j$ is at most $(aa\ldots a)\boldsymbol{c}_j = ac_{1j}+ac_{2j}+\cdots+ac_{nj} = a(c_{1j}+c_{2j}+\cdots+c_{nj}) \leq Ma$ since the column sum of $\boldsymbol{c}_j$ is less than or equal to M. This proves the base case.

Now assume we have shown that every entry of C^k is less than or equal to $M^{k-1}a$. Consider the $(i, j)-$entry of $C^{k+1} = C^kC$. Let $\boldsymbol{r}'_i$ be the i^{th} row of C^k. By what we have already shown we know that every entry of $\boldsymbol{r}'_i$ is non-negative and at most $M^{k-1}a$. Since all the entries of $\boldsymbol{c}_j$ are non-negative it follows that

$$\boldsymbol{r}'_i\boldsymbol{c}_j \leq (M^{k-1}aM^{k-1}a\ldots M^{k-1}a)\boldsymbol{c}_j = M^{k-1}a(c_{1j}+c_{2j}+\cdots+c_{nj}) \leq M^{k-1}a\times M = M^ka$$

since the column sum of $\boldsymbol{c}_j$ is less than or equal to M.

Since $M < 1$ we can find an n such that $M^{n-1}a < \epsilon$. For this n all the entries of C^n (and C^k for $k \geq n$) are less than ϵ. $\square$

When the **consumption matrix** C satisfies the condition of **Theorem** (3.7.5) and one approximates $(I_n - C)^{-1}$ by $I_n + C + C^2 + \ldots C^{n-1}$ there is economic significance to the products $C^k\boldsymbol{D}$. If the CEOs of this economy were simple minded, knowing that the final external demand is $\boldsymbol{D}$ they would plan to produce just this much. However, to do so, requires an intermediate demand of $C\boldsymbol{D}$ which would then have to be planned for. But this intermediate demand requires a secondary intermediate demand of $C(C\boldsymbol{D}) = C^2\boldsymbol{D}$ and so on. This illustrates what is known as the multiplier effect.

Exercises

In exercises 1 - 3 verify that the given matrix T is a **stochastic matrix**. For the given vector $\boldsymbol{x}_0$ compute $\boldsymbol{x}_1, \boldsymbol{x}_2$ where $\boldsymbol{x}_k = T^k\boldsymbol{x}_0$ and find a **steady state vector** for T.

1. $T = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix}, \boldsymbol{x}_0 = \begin{pmatrix} .6 \\ .4 \end{pmatrix}$

2. $T = \begin{pmatrix} .7 & .3 & .2 \\ .2 & .5 & .2 \\ .1 & .2 & .6 \end{pmatrix}, \boldsymbol{x}_0 = \begin{pmatrix} .4 \\ .3 \\ .3 \end{pmatrix}$

3. $T = \begin{pmatrix} .8 & .3 & 0 \\ .1 & .6 & .3 \\ .1 & .1 & .7 \end{pmatrix}, \boldsymbol{x}_0 = \begin{pmatrix} .7 \\ .1 \\ .2 \end{pmatrix}$

4. In a particular math class, students will take ten quizzes on which they will be scored well prepared (W), adequately prepared (A) or poorly prepared (P). From the experience of previous years it has been determined that when a student scores W on a quiz the probability that they score W on the succeeding quiz is .8 and the probability that they score A is .2. When they score A on a quiz the probability that they follow this with a W is .3, score A again is .5 and the probability that they decline to P is .2. Finally, a student scoring P on a quiz has a .3 probability of raising this to an A on the succeeding quiz and .7 probability of getting another P.

a) Write down a **transition matrix**.

b) What is the probability that a student was well prepared on quiz 1 and poorly prepared on quiz 3?

c) What is the probability that a student did poorly on quiz 3 and well on quiz 5?

d) If a class has 100 students and on the first quiz 90 were well prepared and 10 did adquately what do you expect the distribution to be on the final quiz?

5. The imaginary country of Oceana consists of two lush islands, a large island (L) and a small island (S). The population of the large island is 120,000 and the population of the smaller island is 80,000. Over the course of a year 5% of the inhabitants of the large island move to the small island (and the remainder stay on the large island) while 10% of the inhabitants of the small island move to the larger island.

a) Write down a **transition matrix** which describes this pattern of migration.

		Population from		
		LE	LW	S
	LE	.8	.1	.07
Population to	LW	.15	.85	.03
	S	.05	.05	.9

b) What are the expected populations of the two islands one year from now? In two years? In five years?

c) Over the long term what is the expected population of the two islands?

6. The large island of Oceana has two major residential areas, the East (LE) and the West (LW) with respective populations 100,000 and 20,000. Taking this refinement into consideration the **transition matrix** for migration between the major parts of Oceana is given in the table above.

a) What will be the population of the regions of Oceana 1 year from now? In 2 years? In 5 years?

b) Over the long term what is the expected distribution of the population of Oceana between LE, LW and S?

7. In a particular town the probability that a dry day is followed by another dry day is .75 and the probability that a wet day is followed by another wet day is .45. Determine the expected distribution of dry and wet days over 10 years.

8. Three similar magazines P, S, T are competing for subscribers. Marketing research has found that over any given year 80% of subscribers to P resubscribe, 5% drop their subscription to P and subscribe to S and another 15% go to T. Likewise, over the year 75% of subscribers to S are satisified while 15% go to P and 10% go to T. Finally. 90% of T's readers continue their subscriptions and the remaining subscribers drop their subscription in favor of P.

a) Construct a **stochastic matrix** which describes the renewal habits of the subscribers to these magazines.

b) If initially each magazine has 500,000 subscribers what is the long term expected distribution of these 1.5 million readers?

In exercises 9 - 12 assume the given matrix S is the **population projection matrix** of a species. Determine if the matrix has a **stationary vector**.

9. $S = \begin{pmatrix} 0 & 2 & 2 \\ \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$

10. $S = \begin{pmatrix} 0 & 3 & 4 \\ \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$

11. $S = \begin{pmatrix} 0 & 2 & 4 \\ \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$

12. $S = \begin{pmatrix} 0 & 2 & 3 & 3 \\ \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \end{pmatrix}$

13. A population of squirrels has a lifespan of four years meaning that all the squirrels die off by age four and the entire population is distributed among the age classes: 0 - 1 year, 1 - 2 years, 2 - 3 years and 3-4 years The squirrels become mature and begin breeding after one year of age. In the first year a female produces 2 female in the second year an average of 3 and then 1.5 in their final year. On average, only one third of female offspring live to year 1, one half of those age 1 survive to age 2 and one quarter of those of age 2 live to age 3.

a) Write down the **population projection matrix** for this species.

b) If the initial population of squirrels of squirrels is $\boldsymbol{p}(0) = \begin{pmatrix} 600 \\ 500 \\ 300 \\ 200 \end{pmatrix}$ determine the population in years 1, 2, 5 and 10.

c) Will this population eventually become steady, grow in numbers or decline and ultimately become extinct.

14. The sage grouse is a chickenlike bird found in Western North America. Assume that it has a lifespan of three years with age classes 0 - 1, 1-2 and 2 - 3. The sage grouse becomes productive as a yearling and in the first year produces an average 2.5 eggs. In age class 1 - 2 a female typically produces 4.4 eggs and in class 2 - 3 lays an average of 3.6 eggs. Only 20% of eggs are hatched and survive to age one, while 70% of birds in age class 1 -2 survive to the last class.

a) Write down the **population projection matrix** for this species.

b) If the total population is 150,000, equally divided among the classes, determine the population distribution in years 1, 2, 5 and 10.

c) Will this population eventually become steady, grow in numbers or decline and ultimately become extinct.

15. The life of frogs is divided into three stages each of duration one month: tadpoles, juveniles and adult frogs. A scientist followed 2000 newly laid frog eggs and found that 1500 were eaten by fish. By the next month the remaining eggs hatched but of these 500, 350 were also eaten by predators. The remaining ones became mature adults and lived a final month before dying. During that month each female found a mate and each such pair produced 100 eggs.

a) Assume there were equally many males and females at each stage (sex ratio is 50/50). Write down a **Leslie matrix** that models this population.

b) Assume that these observations are typical of the birth and survival rates for this species of frogs and that at some point there are 100 eggs, 100 tadpoles and 100 adults. Determine the population distribution over the next 5 periods.

c) Will this population eventually become steady, grow in numbers or decline and ultimately become extinct.

16. A species of chinook salmon live for four years and are divided into four age classes: 0 - 1, 1 - 2, 2 - 3 and 3 - 4. The respective survival rates are 8%, 75% and 75%. In year four each salmon produces 40 eggs which hatch.

a) Write down the **Leslie matrix** which will model this population.

b) If there is a population in a given year the population vector is $\boldsymbol{p}(0) = \begin{pmatrix} 4000 \\ 3000 \\ 2000 \\ 1000 \end{pmatrix}$ then determine the population in one year, two years, five years and ten years.

c) Will this population eventually become steady, grow in numbers or decline and ultimately become extinct.

In exercises 17 - 20 verify that the given matrix C could be the **consumption matrix** of a **closed economy** and find a **production vector** $\boldsymbol{P}$ such that $C\boldsymbol{P} = \boldsymbol{P}$.

17. $C = \begin{pmatrix} .85 & .30 \\ .15 & .70 \end{pmatrix}$

18. $C = \begin{pmatrix} .60 & .50 & .25 \\ .30 & .40 & .15 \\ .10 & .10 & .60 \end{pmatrix}$

19. $C = \begin{pmatrix} .50 & .25 & .15 \\ .20 & .60 & .35 \\ .30 & .15 & .50 \end{pmatrix}$

20. The national economy of Oceana has three independent sectors: Manufacturing (M), Agriculture (A) and Services (S). Each 100USD produced by M requires 65USD

of M, 25USD of A and 25USD of S. Each 100USD of A requires 25USD of M, 45 of A and 25USD of S. Finally, each 100USD produced by S requires 10USD of M, 15USD of A and 30USD of S. Write down the **consumption matrix** for this economy and determine if the economy is **closed** or **open**.

In exercises 21 - 24 for the given consumption matrix C for and open economy the given demand vector $\boldsymbol{D}$ find the production vector $\boldsymbol{P}$ such that $\boldsymbol{P} = C\boldsymbol{P} + \boldsymbol{D}$.

21. $C = \begin{pmatrix} .6 & .2 \\ .3 & .6 \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} 100 \\ 300 \end{pmatrix}$

22. $C = \begin{pmatrix} .5 & .3 \\ .4 & .6 \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} 200 \\ 300 \end{pmatrix}$

23. $C = \begin{pmatrix} .5 & .1 & .2 \\ .2 & .6 & .2 \\ .1 & .1 & .4 \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} 120 \\ 90 \\ 98 \end{pmatrix}$

24. $C = \begin{pmatrix} .4 & .1 & .2 \\ .2 & .5 & .2 \\ .1 & .2 & .5 \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} 80 \\ 120 \\ 48 \end{pmatrix}$

Challenge Exercises (Problems)

1. Let $\boldsymbol{p}$ be a **probability vector** and T a **stochastic matrix**. Prove that $T\boldsymbol{p}$ is a probability vector.

2. Let T be a **stochastic matrix**. Prove for every natural number k that T^k is a stochastic matrix.

3. Let C be a matrix all of whose column sums are one. Prove that $null(I_n - C) \neq \{\mathbf{0}_n\}$.

4. Let C be a 2×2 **consumption matrix** and assume the column sums of C are strictly less than one. Prove that C is **productive**.

Chapter 4

Determinants

4.1. Introduction to Determinants

We introduce the definition of the determinant of a square $(n \times n)$ matrix for small values of n and then give a recursive definition.

Am I Ready for This Material

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

Our introduction to determinants is self-contained and consequently in this section we will not make reference to any concepts previously introduced except **square matrices**. The exercises do involve **elementary row operations** and **elementary matrices**.

New Concepts

The **determinant of a 2 x 2 matrix**

The **determinant of a 3 x 3 matrix**

The **determinant of a 4 x 4 matrix**

The **(i,j)-submatrix of an n x n matrix**

The **(i,j)-minor of an n x n matrix**

The **(i,j)-cofactor of an n x n matrix**

The **determinant of an n x n matrix**

Theory (Why It Works)

Recall that when we first investigated the concept of **invertibility** of a **square matrix**, in the simplest case, that of a 2×2 matrix, we proved in **Theorem** (3.4.3) if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then A is invertible if and only if $ad - bc \neq 0$.

This function of the entries of a 2×2 matrix is the simplest instance of a determinant which we will ultimately define recursively.

Definition 4.1. Determinant of a 2 x 2 matrix

For a 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ the ***determinant*** of A, denoted by $det(A)$, is defined as $a_{11}a_{22} - a_{12}a_{21}$.

Example 4.1.1. 1. $det(\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}) = 2 \times 5 - 3 \times 4 = 10 - 12 = -2.$

2. $det(\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}) = 1 \times 6 - 2 \times 3 = 6 - 6 = 0.$

3. $det(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}) = 0 \times 0 - 1 \times 1 = 0 - 1 = -1.$

4. $det(\begin{pmatrix} 1 & 5 \\ 0 & 1 \end{pmatrix}) = 1 \times 1 - 5 \times 0 = 1 - 0 = 1.$

5. $det(\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}) = 1 \times 1 - 0 \times (-3) = 1 - 0 = 1.$

6. $det(\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}) = 3 \times 1 - 0 \times 0 = 3 - 0 = 3.$

Definition 4.2. Determinant of a 3 x 3 matrix

Let A be a 3×3 matrix, $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$.

The ***determinant of*** A is defined as $det(A) =$

$$a_{11}det(\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}) - a_{12}det(\begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix}) + a_{13}det(\begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

Remark 4.1. Making use of the definition of the **determinant of a 2 x 2 matrix** it is easy to see that the determinant of the 3 x 3 matrix A is equal to

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}.$$

Notice that each term has the form $a_{1i}a_{2j}a_{3k}$ where the $\{i, j, k\}$ is a permutation (a listing) of 1,2,3 of which there are six possibilities:

$$123, 231, 312, 132, 321, 213 \tag{4.1}$$

You can predict the sign of the term by whether the permutation cyclically permutes 1,2,3 (+ sign) (the first three permutations in (4.1)) or leaves exactly one of 1,2,3 fixed and exchanges the other two (- sign).

There is an easy way to remember this: make the 3×5 matrix from A by augmenting A with its first and second columns so that you obtain the matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{pmatrix}. \tag{4.2}$$

Now start at the entry (1,1) and go down and to the right and multiply the three entries, then move to the right to the (1,2) entry and do the same and again to the right to the

(1,3) entry. These are the positive terms. If you start at the (3,1) entry and go up and to the right and multiply the three entries, then move to the (3,2) entry and then the (3,3) entry in this way you get the negative terms.

We do an example.

Example 4.1.2. Find the [determinant of the 3 x 3 matrix](#)

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 4 & 9 & 16 \end{pmatrix}.$$

we form the matrix $\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 2 & 3 \\ 4 & 9 & 16 & 4 & 9 \end{pmatrix}$ and compute the determinant as just described:

$$(1)(3)(16) + (1)(4)(4) + (1)(2)(9) - (4)(3)(1) - (9)(4)(1) - (16)(2)(1)$$

$$= 48 + 16 + 18 - 12 - 36 - 32$$

$$= 2.$$

Definition 4.3. Determinant of a 4 x 4 matrix

Suppose A is the matrix $\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$. Then the ***determinant*** of A, denote by $det(A)$, is defined to be

$$a_{11} det \begin{pmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{pmatrix} - a_{12} det \begin{pmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{pmatrix} +$$

$$a_{13} det \begin{pmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{pmatrix} - a_{14} det \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{pmatrix}.$$

Example 4.1.3. Find the determinant of the 4×4 matrix in which the (i,j)-entry is $|i-j|$. The matrix is $\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}$.

Then the determinant is

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

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

$$= 0 \times 4 - 1 \times 6 + 2 \times 0 - 3 \times 2$$

$$= 0 - 6 + 0 - 6 = -12.$$

Example 4.1.4. Let D be the matrix whose $(i,j)-$entry is j^{i-1} so that $D = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}$.

Find the **determinant** of D.

$$det(D) = (1)det \begin{pmatrix} 2 & 3 & 4 \\ 4 & 9 & 16 \\ 8 & 27 & 64 \end{pmatrix} - (1)det \begin{pmatrix} 1 & 3 & 4 \\ 1 & 9 & 16 \\ 1 & 27 & 64 \end{pmatrix}$$

$$+(1)det \begin{pmatrix} 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 8 & 64 \end{pmatrix} - (1)det \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix}$$

$$det \begin{pmatrix} 2 & 3 & 4 \\ 4 & 9 & 16 \\ 8 & 27 & 64 \end{pmatrix} =$$

$$2 \times 9 \times 64 + 3 \times 16 \times 8 + 4 \times 4 \times 27 - 8 \times 9 \times 4 - 27 \times 16 \times 2 - 64 \times 4 \times 3 = 48$$

$$det \begin{pmatrix} 1 & 3 & 4 \\ 1 & 9 & 16 \\ 1 & 27 & 64 \end{pmatrix} =$$

$$1 \times 9 \times 64 + 3 \times 16 \times 1 + 4 \times 1 \times 27 - 1 \times 9 \times 4 - 27 \times 16 \times 1 - 64 \times 1 \times 3 = 72$$

$$det \begin{pmatrix} 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 8 & 64 \end{pmatrix} =$$

$$1 \times 4 \times 64 + 2 \times 16 \times 1 + 4 \times 1 \times 8 - 1 \times 4 \times 4 - 8 \times 16 \times 1 - 64 \times 1 \times 2 = 48$$

$$det \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix} =$$

$$1 \times 4 \times 27 + 2 \times 9 \times 1 + 3 \times 1 \times 8 - 1 \times 4 \times 3 - 8 \times 9 \times 1 - 27 \times 1 \times 2 = 12$$

So, $det(D) = 48 - 72 + 48 - 12 = 12$.

(Check this with a calculator.)

We could go on like this but we give a more uniform formulation. We introduce several new concepts leading to a **recursive definition** of the determinant.

Definition 4.4. Let A be an $n \times n$ matrix and i, j a pair of indices, $1 \leq i, j \leq n$. Then the $(i, j)-$***submatrix of*** A is the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i^{th} row and the j^{th} column. We will denote this by A_{ij}.

Example 4.1.5. Find the (1,1)-submatrix of $A = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 9 & 16 \\ 8 & 27 & 64 \end{pmatrix}$.

The (1,1)-submatrix of A is $\begin{pmatrix} 9 & 16 \\ 27 & 64 \end{pmatrix}$.

Find the (1,3)-submatrix of A.

The (1,3)-submatrix of A is $\begin{pmatrix} 4 & 9 \\ 8 & 27 \end{pmatrix}$.

Compute the (3,2)-submatrix of A.

The (3,2)-submatrix of A is $\begin{pmatrix} 2 & 4 \\ 4 & 16 \end{pmatrix}$.

Now assume that we have defined the determinant of an $(n-1) \times (n-1)$ matrix.

Definition 4.5. Suppose that A is an $n \times n$. matrix. The $(i,j)-$***minor of*** A is defined to be the determinant of the $(i,j)-$submatrix of A , which presumably we know how to compute since it is an $(n-1) \times (n-1)$ matrix. We denote this by $M_{ij}(A)$. So, $M_{ij}(A) = det(A_{ij})$.

Example 4.1.6. Compute the (2,4)-minor of the matrix $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}$.

The **(2,4) submatrix** of A is $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix}$. The **determinant of this 3 x 3 matrix** is

$$(1)(4)(27) + (1)(9)(1) + (1)(1)(8) - (1)(4)(1) - (8)(9)(1) - (27)(1)(1) = 22$$

Definition 4.6. Let A be an $n \times n$ matrix. Then the $(i,j)-$**cofactor** of A is defined to be $(-1)^{i+j} M_{ij}(A) = (-1)^{i+j} det(A_{ij})$. That is, it is the **(i,j)-minor** times the sign $(-1)^{i+j}$. We denote the $(i,j)-$cofactor by $C_{ij}(A)$.

Example 4.1.7. Determine the (4,3)-cofactor of the matrix A of **Example** (4.1.6).

The **(4,3)-submatrix** of A is $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{pmatrix}$. The (4,3)-minor of A is

$$(1)(2)(16) + (1)(4)(1) + (1)(1)(4) - (1)(2)(1) - (4)(4)(1) - (16)(1)(1) = 6.$$

Then the (4,3)-cofactor is $(-1)^{4+3}(6) = -6$.

Finally, we can give the **recursive definition** of the determinant:

Definition 4.7. Let $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$. Then the ***determinant of*** A is defined to be

$$a_{11}C_{11}(A) + a_{12}C_{12}(A) + \cdots + a_{1n}C_{1n}(A). \tag{4.3}$$

In words: we multiply each entry of the first row by the corresponding **cofactor** and add these up. The result is the determinant of A.

Note that this is the definition of the determinant. It is called the **cofactor expansion in the first row**.

Remark 4.2. As we shall show there is nothing special about the first row: actually any row yields the same number (we shall show this and not assume it). Once we have established this then we will also be able to compute the determinant as

$$a_{i1}C_{i1}(A) + a_{i2}C_{i2}(A) + \cdots + a_{in}C_{in}(A) \tag{4.4}$$

for any i where $1 \leq i \leq n$.

This is referred to as the ***cofactor expansion*** in the i^{th} row.

In investigating properties of the **determinant** we will also show for a square matrix A the determinant of A and the determinant of the **transpose** of A are equal. Since the determinant of A^{Tr} can be computed by the cofactor expansion in any of its rows and the rows of A^{Tr} are the columns of A it will then follow that the determinant of A can be computed by a ***cofactor expansion of one of its columns***:

$$a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \tag{4.5}$$

Example 4.1.8. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 4 & 9 & 16 \end{pmatrix}$ a matrix whose determinant we previously computed in **Example** (4.1.2) by the cofactor expansion in the first row. Calculate the cofactor expansion in the remaining rows and in the columns.

Expansion in the second row:

$$2 \times (-1)^{2+1} det \begin{pmatrix} 1 & 1 \\ 9 & 16 \end{pmatrix} + 3 \times (-1)^{2+2} det \begin{pmatrix} 1 & 1 \\ 4 & 16 \end{pmatrix} + 4 \times (-1)^{2+3} det \begin{pmatrix} 1 & 1 \\ 4 & 9 \end{pmatrix} =$$

$$2 \times (-7) + 3 \times (12) + 4 \times (-5) = -14 + 36 - 20 = 2.$$

Expansion in the third row gives:

$$4 \times (-1)^{3+1} det \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix} + 9 \times (-1)^{3+2} det \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix} + 16 \times (-1)^{3+3} det \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} =$$

$$4 \times 1 + 9 \times (-2) + 16 \times 1 = 4 - 18 + 16 = 2.$$

Expansion in the first column:

$$1 \times (-1)^{1+1} det \begin{pmatrix} 3 & 4 \\ 9 & 16 \end{pmatrix} + 2 \times (-1)^{1+2} det \begin{pmatrix} 1 & 1 \\ 9 & 16 \end{pmatrix} + 4 \times (-1)^{1+3} det \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix} =$$

$$1 \times 12 - 2 \times 7 + 4 \times 1 = 2$$

Expansion in the second column:

$$1 \times (-1)^{1+2} det \begin{pmatrix} 2 & 4 \\ 4 & 16 \end{pmatrix} + 3 \times (-1)^{2+2} det \begin{pmatrix} 1 & 1 \\ 4 & 16 \end{pmatrix} + 9 \times (-1)^{3+2} det \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix} =$$

$$1 \times (-16) + 3 \times 12 + 9 \times (-2) = 2$$

Expansion in the third column:

$$1 \times (-1)^{1+3} det \begin{pmatrix} 2 & 3 \\ 4 & 9 \end{pmatrix} + 4 \times (-1)^{2+3} det \begin{pmatrix} 1 & 1 \\ 4 & 16 \end{pmatrix} + 16 \times (-1)^{3+3} det \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} =$$

$$1 \times 6 + 4 \times (-5) + 16 \times 1 = 2$$

Remark 4.3. **Properties of Determinants**

While the **recursive definition** of the **determinant** can be used to compute, it is not particularly efficient. If the matrix is $n \times n$ then the determinant will involve the addition and subtraction of $n! = n \times (n-1) \times \cdots \times 2 \times 1$ terms, with each term being the product of n numbers, one from each row and column, and consequently involving $n-1$ multiplications. That's a lot of work. In the next section we develop some properties of the determinant that will allow us to compute much more efficiently.

What You Can Now Do

1. Write down the **(i,j)-submatrix of a matrix**.
2. Compute the **(i,j)-minor of a matrix**.
3. Compute the **(i,j)-cofactor of a matrix**.
4. Compute the determinant of a matrix using the **cofactor expansion in the first row**.

Method (How To Do It)

Method 4.1.1. Write down the **(i,j)-submatrix of a matrix**.

Simply cross out the i^{th} row and the j^{th} column and copy the remaining matrix.

Example 4.1.9. Write down the (2,2)- and (1,3)- **submatrix** of the matrix

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

The **(2,2)-submatrix** is

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 4 & -8 \\ -1 & 7 & -3 \end{pmatrix}$$

while the **(1,3)-submatrix** is

$$\begin{pmatrix} -2 & 4 & 6 \\ 0 & 2 & -8 \\ -1 & 5 & -3 \end{pmatrix}$$

Method 4.1.2. Compute the **(i,j)-minor of a matrix**.

First write down the **(i,j)-submatrix**. Then take its **determinant**.

Example 4.1.10. Let $A = \begin{pmatrix} 4 & -3 & 1 \\ 2 & -2 & 5 \\ -7 & 2 & 4 \end{pmatrix}$.

a) Compute the **(1,1)-minor** of A.

$$det \begin{pmatrix} -2 & 5 \\ 2 & 4 \end{pmatrix} = (-2)(4) - (5)(2) = -8 - 10 = -18$$

b) Compute the **(3,2)-minor** of A.

$$det \begin{pmatrix} 4 & 1 \\ 2 & 5 \end{pmatrix} = (4)(5) - (1)(2) = 20 - 2 = 18$$

Example 4.1.11. Let $A = \begin{pmatrix} 1 & -1 & 2 & 2 \\ 0 & 3 & -1 & 1 \\ 2 & -1 & 4 & 3 \\ 1 & 2 & -3 & 4 \end{pmatrix}$.

a) Compute the **(1,4)-minor** of A.

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

$$0 \times (-1) \times (-3) + 3 \times 4 \times 1 + (-1) \times 2 \times 2$$

$$-[1 \times (-1) \times (-1)] - [2 \times 4 \times 0] - [(-3) \times 2 \times 3] =$$

$$0 + 12 + (-4) - 1 - 0 - (-18) = 25.$$

b) Compute the **(3,3)-minor** of A.

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

$$1 \times 3 \times 4 + (-1) \times 1 \times 1 + 2 \times 0 \times 2$$

$$-[1 \times 3 \times 2] - [2 \times 1 \times 1] - [4 \times 0 \times (-1)] =$$

$$12 + (-1) + 0 - 6 - 2 - 0 = 3.$$

Method 4.1.3. Compute the **(i,j)-cofactor of a matrix**.

Compute the **(i,j)-minor** and multiply by $(-1)^{i+j}$.

Example 4.1.12. Let $A = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 1 \\ 1 & 9 & 25 \end{pmatrix}$.

a) Compute the **(1,2) cofactor** of A.

$$(-1)^{1+2} det \begin{pmatrix} 1 & 1 \\ 1 & 25 \end{pmatrix} = -[(1)(25) - (1)(1)] = -24$$

b) Compute the **(3,3) cofactor** of A.

$$(-1)^{3+3} det \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix} = (1)(1) - (1)(3) = -2$$

Method 4.1.4. Compute the determinant of a matrix using the **cofactor expansion in the first row**.

Multiply each entry of the first row by the corresponding cofactor and add up.

Example 4.1.13. Let $A = \begin{pmatrix} 1 & -2 & 3 & -4 \\ 2 & -1 & 2 & -3 \\ 3 & 1 & 4 & 9 \\ 4 & -1 & 8 & 9 \end{pmatrix}$. Compute the determinant of A using the **cofactor expansion in the first row**.

For a general 4×4 matrix $\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$ this is

$$a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} + a_{14}C_{14}$$

For the given matrix the **cofactor expansion in the first row** is

$$1 \times (-1)^{1+1} det \begin{pmatrix} -1 & 2 & -3 \\ 1 & 4 & 9 \\ -1 & 8 & 9 \end{pmatrix} + (-2) \times (-1)^{1+2} det \begin{pmatrix} 2 & 2 & -3 \\ 3 & 4 & 9 \\ 4 & 8 & 9 \end{pmatrix} +$$

$$3 \times (-1)^{1+3} det \begin{pmatrix} 2 & -1 & -3 \\ 3 & 1 & 9 \\ 4 & -1 & 9 \end{pmatrix} + (-4) \times (-1)^{1+4} det \begin{pmatrix} 2 & -1 & 2 \\ 3 & 1 & 4 \\ 4 & -1 & 8 \end{pmatrix}.$$

We use the **cofactor expansion in the first row** to compute the determinants of the four matrices

$$\begin{pmatrix} -1 & 2 & -3 \\ 1 & 4 & 9 \\ -1 & 8 & 9 \end{pmatrix}, \begin{pmatrix} 2 & 2 & -3 \\ 3 & 4 & 9 \\ 4 & 8 & 9 \end{pmatrix}, \begin{pmatrix} 2 & -1 & -3 \\ 3 & 1 & 9 \\ 4 & -1 & 9 \end{pmatrix}, \begin{pmatrix} 2 & -1 & 2 \\ 3 & 1 & 4 \\ 4 & -1 & 8 \end{pmatrix}$$

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

$$(-1)\times(-1)^{1+1} det \begin{pmatrix} 4 & 9 \\ 8 & 9 \end{pmatrix} + 2\times(-1)^{1+2} det \begin{pmatrix} 1 & 9 \\ -1 & 9 \end{pmatrix} + (-3)\times(-1)^{1+3} det \begin{pmatrix} 1 & 4 \\ -1 & 8 \end{pmatrix} =$$

$$(-1)[36 - 72] + 2(-1)[9 - (-9)] + (-3)[8 - (-4)] = 36 - 36 - 36 = -36.$$

$$det \begin{pmatrix} 2 & 2 & -3 \\ 3 & 4 & 9 \\ 4 & 8 & 9 \end{pmatrix} =$$

$$2 \times (-1)^{1+1} det \begin{pmatrix} 4 & 9 \\ 8 & 9 \end{pmatrix} + 2 \times (-1)^{1+2} det \begin{pmatrix} 3 & 9 \\ 4 & 9 \end{pmatrix} + (-3) \times (-1)^{1+3} det \begin{pmatrix} 3 & 4 \\ 4 & 8 \end{pmatrix} =$$

$$2 \times [36 - 72] + 2(-1)[27 - 36] + (-3)[24 - 16] = -72 + 18 - 24 = -78$$

$$det \begin{pmatrix} 2 & -1 & -3 \\ 3 & 1 & 9 \\ 4 & -1 & 9 \end{pmatrix} =$$

$$2\times(-1)^{1+1}det \begin{pmatrix} 1 & 9 \\ -1 & 9 \end{pmatrix} + (-1)\times(-1)^{1+2}det \begin{pmatrix} 3 & 9 \\ 4 & 9 \end{pmatrix} + (-3)\times(-1)^{1+3}det \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix} =$$

$$2[9-(-9)] + (-1)(-1)[27-36] + (-3)[(-3)-4] = 36 - 9 + 21 = 48$$

$$det \begin{pmatrix} 2 & -1 & 2 \\ 3 & 1 & 4 \\ 4 & -1 & 8 \end{pmatrix} =$$

$$2\times(-1)^{1+1}det \begin{pmatrix} 1 & 4 \\ -1 & 8 \end{pmatrix} + (-1)\times(-1)^{1+2}det \begin{pmatrix} 3 & 4 \\ 4 & 8 \end{pmatrix} + 2\times(-1)^{1+3}det \begin{pmatrix} 3 & 1 \\ 4 & -1 \end{pmatrix} =$$

$$2[8-(-4)] + (-1)(-1)[24-16] + 2[(-3)-4] = 24 + 8 - 14 = 18$$

Now we can substitute these in the **cofactor expansion in the first row** for A to compute its determinant:

$$det(A) = 1 \times (-36) + (-2) \times 78 + 3 \times 48 + (-4) \times (-18) = 24.$$

Exercises

For the matrices in 1-4 write down the (1,3)-, (2,2)- and (3,2)- **submatrix**. See **Method** (4.1.1).

1. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

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

3. $\begin{pmatrix} -4 & 1 & 2 \\ 3 & -2 & 0 \\ 5 & 7 & -6 \end{pmatrix}$

4. $\begin{pmatrix} 2 & -3 & 5 & 4 \\ -6 & 1 & 4 & 3 \\ -7 & 5 & 2 & 2 \\ 1 & 3 & -2 & 6 \end{pmatrix}$

For each of the matrices in 5- 8 write down the (1,2)-, (2,3)- and (3,1)- **minor**. See **Method** (4.1.2).

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

6. $\begin{pmatrix} 3 & 7 & -4 \\ 2 & -3 & 5 \\ 0 & 1 & -2 \end{pmatrix}$

7. $\begin{pmatrix} -8 & 7 & -6 \\ 4 & 5 & 6 \\ 3 & -2 & 1 \end{pmatrix}$

8. $\begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 8 \\ -2 & 3 & -4 & 5 \end{pmatrix}$

In exercises 9 and 10 compute the (1,1)-, (1,2)-, (1,3)-, (2,1)- and (3,1)- **cofactors**. See **Method** (4.1.3).

9. $\begin{pmatrix} 1 & 3 & 4 \\ -1 & 2 & 5 \\ 2 & 1 & -1 \end{pmatrix}$

10. $\begin{pmatrix} 4 & 2 & 1 \\ 3 & 1 & 0 \\ 5 & 3 & 1 \end{pmatrix}$

Compute the determinants of the following matrices using the **cofactor expansion in the first row**. See **Method** (4.1.4).

11. $\begin{pmatrix} 3 & 2 & 0 \\ 4 & 5 & -1 \\ 1 & 3 & -1 \end{pmatrix}$

12. $\begin{pmatrix} 0 & -1 & 2 \\ 3 & 4 & 7 \\ 6 & 9 & 5 \end{pmatrix}$

In 13 and 14 express (but don't compute the determinants of matrices by the **cofactor expansion in the third row**.

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

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

Compute the determinant of the following matrix using the cofactor expansion of your choice. See **Remark** (4.2).

15. $\begin{pmatrix} 1 & 0 & -1 & 2 \\ 3 & 2 & 0 & -4 \\ -5 & 0 & 2 & 0 \\ 3 & -1 & 1 & -6 \end{pmatrix}$

In 16 - 20 compute the determinants of each group of matrices and explain what conclusions you might draw from these examples. See **Method** (4.1.4).

16. $A; B = (R_1 \to 3R_1)A; C = (R_2 \to 2R_2)A$ where
$A = \begin{pmatrix} 1 & 3 & -2 \\ 5 & -1 & 7 \\ 3 & 4 & 2 \end{pmatrix}$.

17. $A; B = (R_1 \leftrightarrow R_2)A; C = (R_1 \leftrightarrow R_3)A$ where
$A = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 7 & 4 \\ -2 & 1 & 6 \end{pmatrix}$.

18. $A; A^{Tr}$ where
$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix}$.

19. $A = \begin{pmatrix} 1 & 3 & 5 \\ 2 & -3 & 6 \\ 1 & 3 & 5 \end{pmatrix}, B = \begin{pmatrix} 1 & 4 & 7 \\ 1 & 4 & 7 \\ 2 & -5 & 6 \end{pmatrix}, C = \begin{pmatrix} 3 & 5 & 7 \\ 2 & 1 & 4 \\ 2 & 1 & 4 \end{pmatrix}$.

20. $A; B = [R_2 \to (-2)R_1 + R_2]A; C = [R_3 \to (-3)R_1 + R_3]A$ where
$A = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 7 & -7 \\ 3 & -7 & -8 \end{pmatrix}$.

21. Demonstrate the following identity: $\begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{pmatrix} = (c-b)(c-a)(b-a)$.

See **how to compute the determinant of a 3 x 3 matrix**.

In 22-28 compute the determinants of the given **elementary matrices**. See **Method** (4.1.4).

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

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

24. $\begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

25. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

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

27. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix}$.

28. $\begin{pmatrix} 1 & 0 & -6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

In exercises 29 - 32 answer true or false and give an explanation.

29. If A and B are 2×2 matrices, and $det(A) = det(B) = 0$ then $det(A + B) = 0$.

30. If A is a 2×2 matrix and $det(A) \neq 0$ then A is **invertible**.

31. If all the entries of an $n \times n$ matrix are positive then $det(A) > 0$.

32. If the first row of an $n \times n$ matrix contains all zeros then $det(A) = 0$.

Challenge Exercises (Problelms)

1. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2 matrix with determinant $ad - bc \neq 0$. Let $\overline{v}_1 = A\overline{e}_1, \overline{v}_2 = A\overline{e}_2$. Prove that the area of the parallelgram with these two vectors as two of its sides has area $|det(A)| = |ad - bc|$.

2. Prove if A is an $n \times n$ **lower triangular matrix** with diagonal elements $d_1, d_2, \dots, d_n$ then $det(A) = d_1 d_2 \dots d_n$. Hint: Use **mathematical induction** on n.

3. Assume A is $n \times n$ and every entry in the first column is zero. Prove that $det(A) = 0$.

4. Prove if A is an $n \times n$ **upper triangular matrix** with diagonal elements $d_1, d_2, \dots, d_n$ then $det(A) = d_1 d_2 \dots d_n$. Hint: Use **mathematical induction** on n.

4.2. Properties of Determinants

Several important properties of the **determinant** are proved and these provide a more effective way to compute the determinant of a square matrix.

Am I Ready for This Material

The following are concepts that will be used in this section:

elementary row operations

elementary matrices

transpose of a matrix

linearly dependent and linearly independent sequences

determinant of an $n \times n$ matrix

Quiz

Compute the following determinants:

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

2. $\begin{pmatrix} -10 & -15 \\ 3 & 4 \end{pmatrix}$

3. $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 7 & 5 \\ 2 & 1 & 6 \end{pmatrix}$

4. $\begin{pmatrix} 2 & 7 & 5 \\ 1 & 2 & 3 \\ 2 & 1 & 6 \end{pmatrix}$

5. $\begin{pmatrix} 2 & 0 & 0 & 0 \\ -2 & 2 & 7 & 5 \\ 3 & 1 & 2 & 3 \\ -4 & 2 & 1 & 6 \end{pmatrix}$

Quiz Solutions

New Concepts

There are no new concepts in this section.

Theory (Why It Works)

Recall in Section (4.1) we **recursively defined** the concept of the **determinant**. We began by defining the determinant of a 2×2 matrix as a specific function of it entries (see **the definition of the determinant of a 2 x 2 matrix**). Then, assuming that we know how to define (compute) the determinant of an $(n-1) \times (n-1)$ matrix, we defined the determinant of an $n \times n$ matrix $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$ by the **cofactor expansion in the first row**:

$$det(A) = a_{11}C_{11}(A) + a_{12}C_{12}(A) + \cdots + a_{1n}C_{1n}(A) \tag{4.6}$$

In (4.6) C_{ij} is the **(i,j)-cofactor**, $C_{ij} = (-1)^{ij} det(A_{ij})$ where A_{ij} is the **(i,j)-submatrix** of A, that is, the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i^{th} row and the j^{th} column.

As mentioned at the end of Section (4.1), while this definition of the determinant can theoretically be used to compute, it is not a particularly efficient. If the matrix is $n \times n$ then the determinant will involve the addition and subtraction of $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$ terms, with each term being the product of n numbers, one from each row and column, and consequently involving $n-1$ multiplications. That's a lot of work. We develop below some properties of the determinant that will allow us to compute much more efficiently.

Lemma 4.2.1. *The* **determinant** *of the* **n x n identity matrix, I_n** *is 1.*

Proof. The proof is a straightforward application of **mathematical induction**. We leave the details of the proof as a **challenge exercise**.

Lemma 4.2.2. *Assume $1 \leq i < j \leq n$. Then $det(P_{ij}^n) = -1$.*

Proof. The proof is by **mathematical induction** on $n \geq 2$.

Base case, $n = 2$. There is only one exchange matrix when $n = 2$, namely $P_{12}^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and by the **definition of the determinant of a 2 x 2 matrix** it is clear that $det(P_{12}^2) = -1$.

Inductive case: Assume $n \geq 2$ and for any $n \times n$ exchange matrix $P_{ij}^n, det(P_{ij}^n) = -1$. Now assume that $1 \leq i < j \leq n+1$. Suppose $i > 1$. Then the (1,1)-entry of P_{ij}^{n+1} is 1 and all other entries in the first row are zero. By **cofactor expansion in the first row** it follows that $det(P_{ij}^{n+1})$ is equal to the determinant of the **(1,1)-cofactor** of P_{ij}^{n+1} which is $P_{i-1,j-1}^n$ which, by the inductive hypothesis, has determinant -1.

We may therefore assume that $i = 1$. Assume first that $j = 2$. The only non-zero entry in the first row of P_{12}^{n+1} is in the second column. Moreover, the **(1,2)-cofactor** of P_{12}^{n+1} is the **n x n identity matrix, I_n**. By **cofactor expansion in the first row** it follows that $det(P_{12}^{n+1}) = (-1)^{1+2}det(I_n)$ which is -1 by **Lemma** (4.2.1).

Thus, we may assume that $i = 1$ and $j > 2$. Now the only entry in the first row occurs in the j^{th} column and therefore the only term we get in the **cofactor expansion in the first row** is $1 \times (-1)^{1+j}$ multiplied by the **(1,j)-minor** of $P_{1,j}^{n+1}$. This $n \times n$ matrix has one non-zero entry in its first row occurring in the second column. Thus, its determinant is $1 \times (-1)^{1+2}$ times its **(1,2)-minor**. If $j > 3$ then the same thing occurs again. When we have repeated this $j-2$ times the matrix that remains is the $(n-j+2) \times (n-j+2)$ identity matrix, which by **Lemma** (4.2.1) has determinant 1. It now follows that the determinant of $P_{1,j}^{n+1}$ is

$$(-1)^{1+j}[(-1)^{1+2}]^{j-2} = (-1)^{4j-5} = -1$$

since $4j - 5$ is an odd number. $\square$

Theorem 4.2.3. *Let A be an $n \times n$ matrix. Assume that the matrix B is obtained from A by exchanging the i^{th} and k^{th} rows. Then $det(B) = -det(A)$.*

Remark 4.4. Recall that the application of an **elementary row operation** on a matrix A can be achieved by multiplying the matrix A on the left by the corresponding **elementary matrix**. In this case the **elementary matrix** is P_{ik}^n. We have seen above in **Lemma** (4.2.2) that $det(P_{ik}^n) = -1$ and consequently we are claiming that $det(P_{ik}A) = det(P_{ik})det(A)$.

Before proceeding to the proof we do an example.

Example 4.2.1. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 4 & 9 & 16 \end{pmatrix}, B = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 1 & 1 \\ 4 & 9 & 16 \end{pmatrix}$ so that B is obtained from A by exchanging the first and second rows.

In **Example** (4.1.2) we computed and found that $det(A) = 2$.

We now compute $det(B)$:

$$det(B) = 2 \times 1 \times 16 + 3 \times 1 \times 4 + 4 \times 1 \times 9 - 2 \times 1 \times 9 - 3 \times 1 \times 16 - 4 \times 1 \times 4 =$$

$$32 + 12 + 36 - 18 - 48 - 16 = 80 - 82 = -2.$$

Proof (**Theorem** (4.2.3)) The proof is by **mathematical induction** on n, the size of the **square matrix** A.

Base case, $n = 2$. In this situation there is only one possiblity - that we switch the first and second rows. So if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $det(A) = ad - bc$.

On the other hand, B, the matrix obtained by exchanging the two rows is $\begin{pmatrix} c & d \\ a & b \end{pmatrix}$ which has determinant $cb - ad = -(ab - cd) = -det(A)$ as claimed.

Before going on to the general argument of passing from n to $n + 1$ we rehearse the argument by showing how we pass from the 2×2 case to the 3×3 case.

Let $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. First, let us suppose that we switch the second and third rows. Then $B = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$.

Using the definition of the determinant we expand along the first row to get $det(B) =$

$$a_{11}(-1)^{1+1}det(\begin{pmatrix} a_{32} & a_{33} \\ a_{22} & a_{23} \end{pmatrix}) + a_{12}(-1)^{1+2}det(\begin{pmatrix} a_{31} & a_{33} \\ a_{21} & a_{23} \end{pmatrix}) +$$

$$a_{13}(-1)^{1+3}det(\begin{pmatrix} a_{31} & a_{32} \\ a_{21} & a_{22} \end{pmatrix}).$$

However, by the result for 2×2 matrices we have

$$(-1)^{1+1}det(\begin{pmatrix} a_{32} & a_{33} \\ a_{22} & a_{23} \end{pmatrix}) = -C_{11}(A), (-1)^{1+2}det(\begin{pmatrix} a_{31} & a_{33} \\ a_{21} & a_{23} \end{pmatrix}) = -C_{12}(A),$$

$$(-1)^{1+3}det(\begin{pmatrix} a_{31} & a_{32} \\ a_{21} & a_{22} \end{pmatrix}) = -C_{13}(A).$$

and therefore $det(B) = a_{11}(-C_{11}(A)) + a_{12}(-C_{12}(A)) + a_{13}(-C_{13}(A)) = -det(A)$.

So, at this point we have demonstrated that $det(P_{23}^3 A) = -det(A) = det(P_{23}^3)det(A)$.

Before we proceed to the other cases we make an observation that allows us to reduce to the case of exchanging the first and second rows, namely, $P_{13} = P_{23}P_{12}P_{23}$.

We have already demonstrated that $det(P_{23}A) = -det(A)$. Suppose we also demonstrate that $det(P_{12}A) = -det(A)$.

Then

$$det(P_{13}A) = det(P_{23}P_{12}P_{23}A) = -det(P_{12}P_{23}A) =$$
$$(-1)(-1)det(P_{23}A) = (-1)(-1)(-1)det(A) = -det(A).$$

So we only have to prove the result when we switch the first and second rows.

In this case if $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ then $B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} =$

$\begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Then the **determinant** of A is

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}.$$

The **determinant** of B is

$$b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{11}b_{23}b_{32} - b_{13}b_{22}b_{31} - b_{12}b_{21}b_{33} =$$

$$a_{21}a_{12}a_{33} + a_{22}a_{13}a_{31} + a_{23}a_{11}a_{32} - a_{21}a_{13}a_{32} - a_{23}a_{12}a_{31} - a_{22}a_{11}a_{33} \quad (4.7)$$

When we order the terms in (4.7) so that they are in the form $a_{1i}a_{2j}a_{3k}$ we get the following expression for $det(B)$:

$$a_{12}a_{21}a_{33} + a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} - a_{13}a_{21}a_{32} - a_{12}a_{23}a_{31} - a_{11}a_{22}a_{33}$$

but this is just $-det(A)$ as required.

The General Case

We are now assuming that the result is true for all $(n-1) \times (n-1)$ matrices, that is, if A is an $(n-1) \times (n-1)$ matrix and B is obtained from A by exchanging two distinct rows then $det(B) = -det(A)$. We then have to prove that the result holds for all $n \times n$ matrices.

First consider the case where we exchange the i^{th} and j^{th} rows where $1 < i < j \leq n$. By the **cofactor expansion in the first row** for the determinant of A we get

$$det(A) = (-1)^{1+1}a_{11}det(A_{11}) + (-1)^{1+2}a_{12}det(A_{12}) + \dots (-1)^{1+n}a_{1n}det(A_{1n})$$

On the other hand, the **cofactor expansion in the first row** for the determinant of B yields

$$det(B) = (-1)^{1+1}b_{11}det(B_{11}) + (-1)^{1+2}b_{12}det(B_{12}) + \dots (-1)^{1+n}b_{1n}det(B_{1n}) =$$

$$(-1)^{1+1}a_{11}det(B_{11}) + (-1)^{1+2}a_{12}det(B_{12}) + \dots (-1)^{1+n}a_{1n}det(B_{1n})$$

since the first rows of A and B are the same. Moreover, each of the matrices B_{1k} is obtained from the matrix A_{1k} by exchanging the $i-1$ and $j-1$ rows. Consequently, for each $k, det(B_{1k}) = -det(A_{1k})$ by the inductive hypothesis. Therefore, in this situation $det(B) = -det(A)$.

We remark that this means that we have shown for $1 < i < j \leq n$ that $det(P_{ij}^n A) = -det(A) = det(P_{ij}^n)det(A)$.

Suppose we can establish the result for the case of exchanging the first and second rows, that is, $det(P_{12}^n A) = -det(A) = det(P_{12}^n A)$. We remark that $P_{1j} = P_{12}P_{2j}P_{12}$. It will then follow that

$$det(P_{1j}A) = det(P_{12}P_{2j}P_{12}A) = -det(P_{2j}P_{12}A) = det(P_{12}A) = -det(A)$$

as required. Thus, it remains to demonstrate the induction in the particular case where we exchange the first and second rows.

Let T be the matrix obtained from A by deleting the first and second rows. For distinct natural numbers $j \neq k$ with $1 \leq j, k \leq n$ let T_{jk} denote the matrix obtained from T by deleting the j and k columns.

When we compute the **determinant** of A by cofactors in the first row we get, as before,

$$det(A) = (-1)^{1+1}a_{11}det(A_{11}) + (-1)^{1+2}a_{12}det(A_{12}) + \dots + (-1)^{1+n}a_{1n}det(A_{1n})$$

In turn we compute the **determinant** of the A_{1k} by **cofactor expansion in the first row**. We will then get a sum over all pairs $j \neq k$ $\epsilon_{jk}a_{1j}a_{2k}det(T_{jk})$ where ϵ_{jk} is a sign, $\pm$.

We need to figure out the signs, ϵ_{jk}. These will depend on whether $j < k$ or $k < j$.

Suppose $j < k$. From the expansion in the first row of A we get a contribution $(-1)^{1+j}$. From the expansion of the first row of A_{1j} we get $(-1)^{(2-1)+(k-1)}$ since in A_{1j} the second row of A is now the first row of A_{1j} and the entries of the k^{th} column are now in the $(k-1)^{st}$ column since $j < k$. Thus, in this case $\epsilon_{jk} = (-1)^{1+j+1+k-1} = (-1)^{1+j+k}$.
By similar reasoning, when $k < j$ we have $\epsilon_{jk} = (-1)^{1+j+1+k} = (-1)^{j+k}$.

Now let us compute the **determinant** of B, the matrix obtained from A by exchanging the first and second rows. Since the matrix obtained from B by deleting the first and second rows is also T we now get the sum taken over all distinct $k \neq j$ with $1 \leq k, j \leq n$ of terms $\delta_{kj}b_{1k}b_{2j}det(T_{jk})$, where δ_{kj} is a sign, $\pm$.

By the same analysis as above, if $k < j$ then $\delta_{kj} = (-1)^{1+j+k}$ whereas if $j < k$ then $\delta_{kj} = (-1)^{j+k}$.

Note that this implies that $\delta_{kj} = -\epsilon_{jk}$ in all cases. Since $b_{1k} = a_{2k}$ and $b_{2j} = a_{1j}$ it now follows that $det(B) = -det(A)$ as required. □

As previously mentioned in **Remark** (4.4), **Theorem** (4.2.3) is equivalent to the following:

Theorem 4.2.4. *Let A be an $n \times n$ matrix and $1 \leq i < j \leq n$. Then $det(P_{ij}A) = det(P_{ij})det(A)$.*

A beautiful consequence of **Theorem** (4.2.3) is the following result on the determinant of a matrix which has two equal rows:

Corollary 4.2.5. *If two rows of an $n \times n$ matrix A are equal, then $det(A) = 0$.*

Proof. The reasoning is as follows: If we switch the two identical rows of A to get B then by **Theorem** (4.2.3) $det(B) = -det(A)$. However, since the rows are equal, $B = A$ and consequently, $det(B) = det(A)$. Thus, $det(A) = -det(A)$ and therefore $det(A) = 0$. □

Still another consequence is that a matrix with a row of zeros has **determinant** equal to zero:

Corollary 4.2.6. *Assume the $n \times n$ matrix A has a row of zeros. Then $det(A) = 0$.*

Proof. Assume the i^{th} row of A consists of all zeros. If $i = 1$ then $det(A) = 0$ by the **cofactor expansion in the first row**. Suppose $i > 1$. Let B be the matrix obtained from A by exchanging the first and i^{th} rows. Then the first row of B consists of all zeros and by what we have already established $det(B) = 0$. However, $det(B) = -det(A)$ by **Theorem** (4.2.3) and therefore $det(A) = 0$. $\square$

Theorem (4.2.3) was challenging but as we will see it pays great dividends. By comparison the remaining results will be easy, mainly because we will be making use of this result.

Our next goal is to determine the effect of scaling on the determinant but before proceeding to this result we state a lemma on the **determinant** of a scaling matrix. We leave the proof as a **challenge exercise**.

Lemma 4.2.7. *Let c be a scalar and $D_i^n(c)$ be the $n \times n$ matrix obtained from the* **identity matrix** *by multiplying the i^{th} row by c. Then $det(D_i^n(c)) = c$.*

We now will prove

Theorem 4.2.8. *Let A be an $n \times n$ matrix. Assume that the matrix B is obtained from A by multiplying the i^{th} row by the scalar c. Then $det(B) = c \times det(A)$.*

Remark 4.5. By **Lemma** (4.2.7) the theorem is equivalent to the assertion that for any $n \times n$ matrix A that $det(D_i^n(c)A) = det(D_j^n(c))det(A)$.

Before proceeding to a demonstration we do an example.

Example 4.2.2. Let C be the matrix obtained from the matrix A of **Example** (4.2.1) by multiplying the second row by -3 so that

$$C = \begin{pmatrix} 1 & 1 & 1 \\ -6 & -9 & -12 \\ 4 & 9 & 16 \end{pmatrix}.$$

Then

$$det(C) = 1 \times (-9) \times 16 + 1 \times (-12) \times 4 + 1 \times (-6) \times 9$$

$$-4 \times (-9) \times 1 - 9 \times (-12) \times 1 - 16 \times (-6) \times 1 =$$

$$-144 - 48 - 54 + 108 + 96 + 36 =$$

$$-246 + 240 = -6 = (-3)det(A).$$

Proof. (**Theorem** (4.2.8)) As indicated by **Remark** (4.5) it suffices to show that

$$det(D_i^n(c)A) = cdet(A) = det(D_i^n(c))det(A).$$

Suppose we have proved the result whenever we scale the first row, that is, multiply by an **elementary matrix** of the form $D_1^n(c)$ so that for any $n \times n$ matrix A

$$det(D_1^n(c)A) = cdet(A) = det(D_1(c))det(A)$$

We remark that in general the scaling matrix for the i^{th} row $D_i^n(c)$ is the same as $P_{1i}D_1^n(c)P_{1i}$: For if we apply this to a matrix A it first exchanges the first and i^{th} rows of A so that now the first row of $P_{1i}A$ is the i^{th} row of A. Then it multiplies the first row of this matrix by c and then it exchanges the first and i^{th} rows. So $P_{1i}D_1^n(c)P_{1i} = D_i^n(c)$ as claimed.

It now follows that

$$det(D_i(c)A) = det(P_{1i}D_1^n(c)P_{1i}A) = -det(D_1^n(c)P_{1i}A) =$$

$$-cdet(P_{1i}A) = cdet(A) = det(D_i^n(c))det(A)$$

as required. Thus, we only have to prove the result when we multiply the first row by a scalar c.

Now by **definition of the determinant** of A by the cofactor expansion in the first row we know the determinant of A is given by the sum

$$(-1)^{1+1}a_{11}det(A_{11}) + (-1)^{1+2}a_{12}det(A_{12}) + \cdots + (-1)^{1+n}a_{1n}det(A_{1n})$$

and the determinant of B by the cofactor expansion

$$(-1)^{1+1}b_{11}det(B_{11}) + (-1)^{1+2}b_{12}det(B_{12}) + \cdots + (-1)^{1+n}b_{1n}det(B_{1n})$$

On the other hand, the matrices $B_{1j} = A_{1j}$ for each value of j since A and B are the same except for the first row. Moreover, $b_{1j} = ca_{1j}$ since B is obtained by multiplying the first row of A by c.

Thus the **determinant** of B is

$$(-1)^{1+1}(ca_{11})det(A_{11})+(-1)^{1+2}(ca_{12})det(A_{12})+\cdots+(-1)^{1+n}(ca_{1n})det(A_{1n}) =$$

$$c[(-1)^{1+1}a_{11}det(A_{11})+(-1)^{1+2}a_{12}det(A_{12})+\cdots+(-1)^{1+n}a_{1n}det(A_{1n})] = cdet(A)$$

□

Our next goal is to characterize the effect that an elimination type of **elementary row operation** has on the **determinant** of a **square matrix**. Before we get to that we need to prove a lemma.

Lemma 4.2.9. *Let A, B and C be $n \times n$ matrices. Assume the sequence of rows of A is $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{i-1}, \boldsymbol{r}_i, \boldsymbol{r}_{i+1}, \ldots, \boldsymbol{r}_n)$; the sequence of rows of B is $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{i-1}, \boldsymbol{r}'_i, \boldsymbol{r}_{i+1}, \ldots, \boldsymbol{r}_n)$; and the sequence of rows of C is $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{i-1}, \boldsymbol{r}_i + \boldsymbol{r}'_i, \boldsymbol{r}_{i+1}, \ldots, \boldsymbol{r}_n)$. In other words, A, B, C all have the same rows except the i^{th} row and the i^{th} row of C is the sum of the i^{th} rows of A and B.*

Then $det(C) = det(A) + det(B)$.

Proof. We first consider the case where $i = 1$ so that A, B, C agree in rows $2, 3 \ldots, n$ and the first row of C is the sum of the first row of A and the first row of B. Note that this implies that the **(1,j)-cofactor** of A, B and C are identical. We shall denote these common cofactors by F_{1j}

Now let $\boldsymbol{r}_1 = (a_{11}, a_{12}, \ldots, a_{1n})$, $\boldsymbol{r}'_1 = (a'_{11}, a'_{12}, \ldots, a'_{1n})$ be the first rows of A and B respectively. By our hypothesis, this implies that the first row of C is $(a_{11} + a'_{11}, a_{12} + a'_{12}, \ldots, a_{1n} + a'_{1n})$.

Now by **cofactor expansion in the first row** for the determinant we have:

$$det(A) = a_{11}F_{11} + a_{12}F_{12} + \cdots + a_{1n}F_{1n}$$

$$det(B) = a'_{11}F_{11} + a'_{12}F_{12} + \cdots + a'_{1n}F_{1n}$$

$$det(C) = (a_{11} + a'_{11})F_{11} + (a_{12} + a'_{12})F_{12} + \cdots + (a_{1n} + a'_{1n})F_{1n}.$$

By the distributive property of real numbers we have

$$det(C) = [a_{11}F_{11} + a'_{11}F_{11}] + [a_{12}F_{12} + a'_{12}F_{12}] + \cdots + [a_{1n}F_{1n} + a'_{1n}F_{1n}] =$$

$$[a_{11}F_{11} + a_{12}F_{12} + \cdots + a_{1n}F_{1n}] + [a'_{11}F_{11} + a'_{12}F_{12} + \cdots + a'_{1n}F_{1n}] =$$

$$det(A) + det(B).$$

We now treat the case where $1 < i$. Let $A' = P_{1i}A, B' = P_{1i}B, C' = P_{1i}C$ be the matrices obtained from A, B, C by switching the first and i^{th} rows. Now the matrices A', B' and C' agree in rows $2, 3, \ldots, n$ and the first row of C' is the sum of the first row of A' and the first row of B'. By what we have already shown,

$$det(C') = det(A') + det(B').$$

Since $A' = P_{1i}A, B' = P_{1i}B, C' = P_{1i}C$ we get

$$det(P_{1i}C) = det(P_{1i}A) + det(P_{1i}B). \tag{4.8}$$

By **Theorem** (4.2.4) it follows that $det(A') = det(P_{1i}A) = -det(A), det(B') = det(P_{1i}B) = -det(B), det(C') = det(P_{1i}C) = -det(C)$. Substituting these into (4.8) and dividing by -1 yields the result. □

Recall, we previously defined the elimination type **elementary matrix** $T_{ij}(c)$ to be matrix obtained from the **identity matrix** by adding c times the i^{th} row to the j^{th} row. This has 1's down the main diagonal, an entry c in the $(j, i)-$entry and 0's everywhere else. If we define E_{ij} to be the $n \times n$ matrix with zeros in all positions except the (i, j) which has a one, then $T_{ij}(c) = I_n + cE_{ji}$.

Remark 4.6. The matrix $T_{ij}^n(c)$ has determinant 1. Since $T_{ij}^n(c)$ is triangular this follows from **challenge exercises 4.1.2 and 4.1.3**.

Theorem 4.2.10. *Let A be an $n \times n$ matrix. If B is obtained from A by adding c times the i^{th} row of A to the j^{th} row of A so that $B = T_{ij}^n(c)A$, then $det(B) = det(A) = det(T_{ij}^n(c))det(A)$.*

Proof. The proof is fairly straightforward. Let $A = \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_{j-1} \\ \boldsymbol{r}_j \\ \boldsymbol{r}_{j+1} \\ \vdots \\ \boldsymbol{r}_n \end{pmatrix}, B = \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_{j-1} \\ \boldsymbol{r}_j + c\boldsymbol{r}_i \\ \boldsymbol{r}_{j+1} \\ \vdots \\ \boldsymbol{r}_n \end{pmatrix}$.

Let B' be the matrix with all rows identical with the rows of A except for the j^{th} row, which we take to be the c times the i^{th} row of A : $B' = \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_{j-1} \\ c\boldsymbol{r}_i \\ \boldsymbol{r}_{j+1} \\ \vdots \\ \boldsymbol{r}_n \end{pmatrix}$.

By **Lemma** (4.2.9) it follows that $det(B) = det(A) + det(B')$.

By **Theorem** (4.2.8) $det(B') = c \times det \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_{j-1} \\ \boldsymbol{r}_i \\ \boldsymbol{r}_{j+1} \\ \vdots \\ \boldsymbol{r}_n \end{pmatrix}$. By **Corollary** (4.2.5) $det(B') = 0$ since B' has two identical rows (the i^{th} and j^{th} rows). Thus, $det(B) = det(A)$ as claimed. $\square$

Remark 4.7. Taken together the three results we have proved imply that if A is an $n \times n$ matrix and E is an $n \times n$ **elementary matrix** then $det(EA) = det(E)det(A)$.

Before showing how these theorems are useful for computing the **determinant** of a matrix we prove a few more results.

Theorem 4.2.11. *An $n \times n$ matrix A is* **invertible** *if and only if $det(A) \neq 0$.*

Proof. Let R be the **reduced echelon form** of A. In **Theorem** (3.5.4) we proved that an $n \times n$ matrix A is **invertible** if and only if R is the **identity matrix, I_n**. In the contrary case, when A is not invertible, it follows from **Lemma** (3.5.3) that R has a row of zeros.

In either case we have seen that we can find elementary matrices $E_1, E_2, \dots, E_k$ such that $E_kE_{k-1}\dots E_2E_1A = R$. It follows by the observation made in **Remark** (4.7) that

$$det(E_k)\dots det(E_1)det(A) = det(R).$$

For each $i, det(E_i) \neq 0$ and therefore $det(A) \neq 0$ if and only if $det(R) \neq 0$. But $det(R) \neq 0$ if and only if R is the **identity matrix** if and only if A is an **invertible matrix**. □

Example 4.2.3. Let $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 5 & 8 \\ 2 & 4 & 9 & 0 \\ -5 & -11 & -17 & -12 \end{pmatrix}$. We show that the **transpose** of A, A^{Tr} is not **non-invertible**.

Notice that $A^{Tr}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ so that $null(A^{Tr}) \neq \{\mathbf{0}_n\}$. This implies that that A^{Tr} is **non-invertible**, and, consequenlty, A is not invertible. Thus, it should be the case that $det(A) = 0$. We use **elementary row operations** to compute $det(A)$:

$$det(A) = det(\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 3 & -8 \\ 0 & -1 & -2 & 8 \end{pmatrix}) = det(\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 3 & -8 \\ 0 & 0 & -3 & 8 \end{pmatrix}) =$$

$$det(\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 3 & -8 \\ 0 & 0 & 0 & 0 \end{pmatrix}) = 0$$

since this last matrix has a row of zeros.

As a corollary of **Theorem** (4.2.11) we have the following:

Corollary 4.2.12. *Assume the* **square matrix** A *has a column of zeros. Then* $det(A) = 0$.

Proof. If A has column of zeros then A is **non invertible** and consequently $det(A) = 0$. □

We now illustrate how the results can be used to compute the **determinant** of a **square matrix** A. Basically, it involves applying **elementary row operations** to A in order to obtain an upper triangular matrix, U. By **Challenge Exercise (4.1.3)** the **determinant** of an **upper triangular matrix** is the product of its **diagonal entries**. By keeping track of the **elementary row operations** and using **Theorem** (4.2.3), **Theorem** (4.2.8) and **Theorem** (4.2.10) we can express the **determinant** of U in terms of the determinant of A. We can compute $det(U)$ by taking the product of its diagonal entries and then we can solve for $det(A)$.

Example 4.2.4. Let A be the 4×4 matrix with $(i, j)-$entry j^{i-1}. Such a matrix is known as a *Vandermonde* matrix. So $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{pmatrix}$.

We obtain an **upper triangular matrix** by applying **elementary row operations**:

$$det(A) = det(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 3 & 8 & 15 \\ 0 & 7 & 26 & 63 \end{pmatrix}) = det(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 12 & 42 \end{pmatrix}) =$$

$$det(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 6 \end{pmatrix}) = 2 \times 6 = 12.$$

The equalities are justified as follows: The second matrix is obtained from A by applying three elimination operations $[R_2 \to (-1)R_1 + R_2, R_3 \to (-1)R_1 + R_3, R_4 \to$

$(-1)R_1 + R_4]$. Therefore by the result on elimination operations the determinants are the same. The third matrix is obtained from the second by two elimination operations $[R_3 \to (-3)R_2 + R_3, R_4 \to (-7)R_2 + R_3]$. So the second and third matrices have the same determinant. Finally, the fourth matrix is obtained from the third by a single elimination operation $[R_4 \to (-6)R_3 + R_4]$. The fourth matrix is **upper triangular** with diagonal entries 1, 1, 2 and 6 and has determinant 12.

We continue to prove some basic properties of the **determinant** .

Theorem 4.2.13. *Let A, B be $n \times n$ matrices. Then $det(AB) = det(A) \times det(B)$.*

Proof. First note that if either A or B is **non-invertible** then by **Theorem** (3.4.16) AB is non-invertible. By **Theorem** (4.2.11), $det(AB) = 0$ and one or both $det(A) = 0, det(B) = 0$. In either case, $det(A)det(B) = 0 = det(AB)$.

We may therefore assume that A and B are both invertible. By **Theorem** (3.5.4) there is a sequence, $(E_1, E_2, \dots, E_k)$ of **elementary matrices** such that $A = E_kE_{k-1}\dots E_2E_1$. Then $AB = E_kE_{k-1}\dots E_2E_1B$.

We have also seen for any $n \times n$ matrix C and elementary matrix $E, det(EC) = det(E)det(C)$. Consequently,

$$det(AB) = det(E_kE_{k-1}\dots E_2E_1B) = det(E_k)det(E_{k-1}\dots E_2E_1B) =$$

$$det(E_k)det(E_{k-1})det(E_{k-2}\dots E_2E_1B) = \dots =$$

$$det(E_k)det(E_{k-1})\dots det(E_2)det(E_1)det(B).$$

However, $det(A) = det(E_k)det(E_{k-1})\dots det(E_2)det(E_1)$ and therefore $det(AB) = det(A)det(B)$. $\square$

In our next theorem we prove that a square matrix A and its **transpose**, A^{Tr}, have the same determinant.

Theorem 4.2.14. *Let A be an $n \times n$ matrix. Then $det(A) = det(A^{Tr})$.*

Proof. We first observe that the result is simple for **elementary matrices**: The matrices P_{ij}^n and $D_i^n(c)$ are their own **transposes** and there is nothing to prove. On the other hand, the **transpose** of an elimination type **elementary matrix** is an elimination type elementary matrix. One is **upper triangular** and the other is **lower triangular**. In any case, both have determinant 1.

Also note that a matrix A is **non-invertible** if and only if its **transpose**, A^{Tr} is non-invertible and therefore $det(A) = 0$ if and only if $det(A^{Tr}) = 0$. Consequently in establishing this theorem it remains only to prove it in the case that A is **invertible**.

By **Theorem** (3.5.4), if A is **invertible** then there is a sequence $(E_1, \dots, E_k)$ of **elemen-
tary matrices** such that $A = E_k \dots E_1$. Then by **Theorem** (4.2.13)

$$det(A) = det(E_k) \dots det(E_1).$$

On the other hand, $A^{Tr} = E_1^{Tr} \dots E_k^{Tr}$ and hence

$$det(A^{Tr}) = det(E_1^{tr}) \dots det(E_k^{Tr}).$$

Since as remarked, $det(E) = det(E^{Tr})$ for an **elementary matrix** it now follows that $det(A^{Tr}) = det(A)$. □

Remark 4.8. It is a consequence of **Theorem** (4.2.14) that everything that has been said about rows, e.g. if two rows are exchanged then the resulting matrix has determinant (-1) times the original determinant, applies equally as well to columns.

Theorem 4.2.15. *The cofactor formula for the* **determinant** *works with any row or any column. More specifically, if A is an $n \times n$ matrix with entries a_{ij} and* **cofactors** *$C_{ij}(A)$ then the determinant can be computed by expanding along the i^{th} row:*

$$det(A) = a_{i1}C_{i1}(A) + a_{i2}C_{i2}(A) + \dots a_{in}C_{in}(A)$$

and can be expanded along the j^{th} column:

$$det(A) = a_{1j}C_{1j}(A) + a_{2j}C_{2j}(A) + \dots a_{nj}C_{nj}(A).$$

Proof. This is left as a **challenge exercise**.

Example 4.2.5. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 4 & 7 \\ 4 & 16 & 49 \end{pmatrix}$. Compute the **determinant** of A using **elementary row operations** and by the **cofactor expansion along the second column**.

We can compute the determinant by applying **elementary row operations** until we obtain an **upper triangule matrix**:

$$det(A) = det(\begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 5 \\ 0 & 12 & 45 \end{pmatrix}) = det(\begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 5 \\ 0 & 0 & 15 \end{pmatrix}) = 1 \times 2 \times 15 = 30.$$

The **cofactor expansion along the second column** is

$$(-1) \times 1 \times (98 - 28) + 4 \times (49 - 4) + (-1) \times 16 \times (7 - 2) = -70 + 180 + 80 = 30.$$

What You Can Now Do

Use **elementary row operations** (equivalently, **elementary matrices**) to compute the **determinant** of a **square matrix** A.

Method (How To Do It)

Method 4.2.1. Use **elementary row operations** (equivalently, **elementary matrices**) to compute the **determinant** of a **square matrix** A.

Find a sequence $(E_1, \dots, E_k)$ of **elementary matrices** such that $U = E_k \dots E_1 A$ is **upper triangular**. Set $A_0 = A$ and for $0 \leq i \leq k-1$ set $A_{i+1} = E_{i+1}A_i$. Express $det(A_{i+1})$ in terms of $det(A_i)$ by using the following:

1) If E_{i+1} is an exchange matrix then $det(A_{i+1}) = -det(A_i)$.

2) If E_{i+1} is a scaling matrix with scaling factor c then $det(A_{i+1}) = c \times det(A_i)$.

3) If E_{i+1} is an elimination matrix then $det(A_{i+1}) = det(A_i)$.

Using 1) - 3) we can express $det(U)$ in terms of $det(A)$. On the other hand, if U has **diagonal entries** $d_1, \dots, d_n$ then $det(U) = d_1 \times \cdots \times d_n$. One can then solve for $det(A)$.

Example 4.2.6. Compute the **determinant** of the matrix $\begin{pmatrix}1 & 3 & 2 & -1\\ 3 & 7 & 2 & 3\\ 2 & 8 & 3 & -8\\ 4 & 4 & 5 & 6\end{pmatrix}$

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

$$(-2)det[\begin{pmatrix}1 & 3 & 2 & -1\\ 0 & 1 & 2 & -3\\ 0 & 2 & -1 & -6\\ 0 & -8 & -3 & 10\end{pmatrix}] = (-2)det[\begin{pmatrix}1 & 3 & 2 & -1\\ 0 & 1 & 2 & -3\\ 0 & 0 & -5 & 0\\ 0 & 0 & 13 & -14\end{pmatrix}] =$$

$$(-2)(-5)det[\begin{pmatrix}1 & 3 & 2 & -1\\ 0 & 1 & 2 & -3\\ 0 & 0 & 1 & 0\\ 0 & 0 & 13 & -14\end{pmatrix}] = (-2)(-5)det[\begin{pmatrix}1 & 3 & 2 & -1\\ 0 & 1 & 2 & -3\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -14\end{pmatrix}] =$$

$$(-2)(-5)(-14) = -140.$$

Example 4.2.7. Compute the **determinant** of the matrix $\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 2 & 3 & 4\\ 1 & 4 & 9 & 16\\ 1 & 8 & 27 & 64\end{pmatrix}$

$$det[\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 2 & 3 & 4\\ 1 & 4 & 9 & 16\\ 1 & 8 & 27 & 64\end{pmatrix}] = det[\begin{pmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 2 & 3\\ 0 & 3 & 8 & 15\\ 0 & 7 & 26 & 63\end{pmatrix}] =$$

$$det[\begin{pmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 2 & 3\\ 0 & 0 & 2 & 6\\ 0 & 0 & 12 & 42\end{pmatrix}] = 2det[\begin{pmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 2 & 3\\ 0 & 0 & 1 & 3\\ 0 & 0 & 12 & 42\end{pmatrix}] =$$

$$2det[\begin{pmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 2 & 3\\ 0 & 0 & 1 & 3\\ 0 & 0 & 0 & 6\end{pmatrix}] = 2 \times 6 = 12.$$

Example 4.2.8. Compute the **determinant** of the matrix $\begin{pmatrix} 0 & 3 & -2 & -1 \\ 2 & 4 & 2 & -8 \\ 3 & 10 & -5 & -8 \\ 3 & 3 & 1 & -7 \end{pmatrix}$

$$det[\begin{pmatrix} 0 & 3 & -2 & -1 \\ 2 & 4 & 2 & -8 \\ 3 & 10 & -5 & -8 \\ 3 & 3 & 1 & -7 \end{pmatrix}] = (2)det[\begin{pmatrix} 0 & 3 & -2 & -1 \\ 1 & 2 & 1 & -4 \\ 3 & 10 & -5 & -8 \\ 3 & 3 & 1 & -7 \end{pmatrix}] =$$

$$(2)(-1)det[\begin{pmatrix} 1 & 2 & 1 & -4 \\ 0 & 3 & -2 & -1 \\ 3 & 10 & -5 & -8 \\ 3 & 3 & 1 & -7 \end{pmatrix}] = (2)(-1)det[\begin{pmatrix} 1 & 2 & 1 & -4 \\ 0 & 3 & -2 & -1 \\ 0 & 4 & -8 & 4 \\ 0 & -3 & -2 & 5 \end{pmatrix}] =$$

$$(2)(-1)(4)det[\begin{pmatrix} 1 & 2 & 1 & -4 \\ 0 & 3 & -2 & -1 \\ 0 & 1 & -2 & 1 \\ 0 & -3 & -2 & 5 \end{pmatrix}] =$$

$$(2)(-1)(4)(-1)det[\begin{pmatrix} 1 & 2 & 1 & -4 \\ 0 & 1 & -2 & 1 \\ 0 & 3 & -2 & -1 \\ 0 & -3 & -2 & 5 \end{pmatrix}] =$$

$$(2)(-1)(4)(-1)det[\begin{pmatrix} 1 & 2 & 1 & -4 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 4 & -4 \\ 0 & 0 & -8 & 8 \end{pmatrix}] =$$

$$(2)(-1)4(-1)(4)(-8)det[\begin{pmatrix} 1 & 2 & 1 & -4 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}] = 0$$

The last conclusion follows from **Theorem** (4.2.5)

Example 4.2.9. Compute the **determinant** of the matrix $\begin{pmatrix} 1 & 2 & 4 & 5 & 2 \\ -1 & 1 & 2 & 1 & 1 \\ 3 & 4 & 8 & 11 & 4 \\ 2 & 5 & -1 & 3 & 6 \\ 1 & 6 & 8 & 13 & 6 \end{pmatrix}$

$$det[\begin{pmatrix} 1 & 2 & 4 & 5 & 2 \\ -1 & 1 & 2 & 1 & 1 \\ 3 & 4 & 8 & 11 & 4 \\ 2 & 5 & -1 & 3 & 6 \\ 1 & 6 & 8 & 13 & 6 \end{pmatrix}] = det[\begin{pmatrix} 1 & 2 & 4 & 5 & 2 \\ 0 & 3 & 6 & 6 & 3 \\ 0 & -2 & -4 & -4 & -2 \\ 0 & 1 & -9 & -7 & 2 \\ 0 & 4 & 4 & 8 & 4 \end{pmatrix}] =$$

$$(3)(-2)(4)det[\begin{pmatrix} 1 & 2 & 4 & 5 & 2 \\ 0 & 1 & 2 & 2 & 1 \\ 0 & 1 & 2 & 2 & 1 \\ 0 & 1 & -9 & -7 & 2 \\ 0 & 1 & 1 & 2 & 1 \end{pmatrix}] = 0.$$

Example 4.2.10. Compute the **determinant** of the matrix $\begin{pmatrix} 2 & 3 & 5 & 2 & 2 \\ 0 & 1 & 3 & 2 & 2 \\ 0 & 2 & 9 & 1 & 4 \\ 3 & 2 & 3 & 2 & -1 \\ 0 & 4 & 1 & 3 & 4 \end{pmatrix}$.

By the **cofactor expansion in the first column** the **determinant** is given by

$$det[\begin{pmatrix} 2 & 3 & 5 & 2 & 2 \\ 0 & 1 & 3 & 2 & 2 \\ 0 & 2 & 9 & 1 & 4 \\ 3 & 2 & 3 & 2 & -1 \\ 0 & 4 & 1 & 3 & 4 \end{pmatrix}] = (2)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 2 & 9 & 1 & 4 \\ 2 & 3 & 2 & -1 \\ 4 & 1 & 3 & 4 \end{pmatrix}] - (3)det[\begin{pmatrix} 3 & 5 & 2 & 2 \\ 1 & 3 & 2 & 2 \\ 2 & 9 & 1 & 4 \\ 4 & 1 & 3 & 4 \end{pmatrix}].$$

We now use properties of determinants to find these two 4×4 determinants:

$$det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 2 & 9 & 1 & 4 \\ 2 & 3 & 2 & -1 \\ 4 & 1 & 3 & 4 \end{pmatrix}] = det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 3 & -3 & 0 \\ 0 & -3 & -2 & -5 \\ 0 & -11 & -5 & -4 \end{pmatrix}] =$$

$$(3)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & -3 & -2 & -5 \\ 0 & -11 & -5 & -4 \end{pmatrix}] = (3)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -5 & -5 \\ 0 & 0 & -16 & -4 \end{pmatrix}] =$$

$$(3)(-5)(-4)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 4 & 1 \end{pmatrix}] = (3)(-5)(-4)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -3 \end{pmatrix}] = -180.$$

$$det[\begin{pmatrix} 3 & 5 & 2 & 2 \\ 1 & 3 & 2 & 2 \\ 2 & 9 & 1 & 4 \\ 4 & 1 & 3 & 4 \end{pmatrix}] = (-1)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 3 & 5 & 2 & 2 \\ 2 & 9 & 1 & 4 \\ 4 & 1 & 3 & 4 \end{pmatrix}] =$$

$$(-1)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & -4 & -4 & -4 \\ 0 & 3 & -3 & 0 \\ 0 & -11 & -5 & -4 \end{pmatrix}] = (-1)(-4)(3)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -11 & -5 & -4 \end{pmatrix}] =$$

$$(-1)(-4)(3)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -1 \\ 0 & 0 & 6 & 7 \end{pmatrix}] = (-1)(-4)(3)det[\begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -1 \\ 0 & 0 & 0 & 4 \end{pmatrix}] =$$

$$(-1)(-4)(3)(-2)(4) = -96.$$

Thus,

$$det[\begin{pmatrix} 2 & 3 & 5 & 2 & 2 \\ 0 & 1 & 3 & 2 & 2 \\ 0 & 2 & 9 & 1 & 4 \\ 3 & 2 & 3 & 2 & -1 \\ 0 & 4 & 1 & 3 & 4 \end{pmatrix}] = 2 \times (-180) - 3 \times (-96) = -72.$$

Exercises

In exercises 1 - 23 compute the **determinant** by any method you can. If you make use of properties of determinants state which properties you are using.

1. $\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 1 & \frac{1}{4} & \frac{1}{9} \\ 1 & \frac{1}{8} & \frac{1}{27} \end{pmatrix}$

2. $\begin{pmatrix} 1 & 3 & 2 \\ 3 & 11 & 10 \\ 2 & 2 & 4 \end{pmatrix}$

3. $\begin{pmatrix} 3 & 9 & -6 \\ 5 & 1 & 3 \\ -2 & -6 & 4 \end{pmatrix}$

4. $\begin{pmatrix} 0 & 2 & 6 \\ 3 & 3 & -6 \\ 4 & 7 & 5 \end{pmatrix}$

5. $\begin{pmatrix} 0 & 1 & 3 \\ -1 & 0 & 7 \\ -3 & -7 & 0 \end{pmatrix}$

6. $\begin{pmatrix} 0 & 3 & -8 \\ -3 & 0 & 5 \\ 8 & -5 & 0 \end{pmatrix}$

7. $\begin{pmatrix} -5 & 1 & 1 \\ 1 & -5 & 1 \\ 1 & 1 & -5 \end{pmatrix}$

8. $\begin{pmatrix} -7 & 2 & 2 \\ 2 & -7 & 2 \\ 2 & 2 & -7 \end{pmatrix}$

9. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \\ 1 & 16 & 81 & 256 \end{pmatrix}$

10. $\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ 1 & \frac{1}{4} & \frac{1}{9} & \frac{1}{16} \\ 1 & \frac{1}{8} & \frac{1}{27} & \frac{1}{64} \\ 1 & \frac{1}{16} & \frac{1}{81} & \frac{1}{256} \end{pmatrix}$

11. $\begin{pmatrix} 3 & -5 & 2 & 7 \\ 0 & 1 & 2 & 4 \\ 0 & 2 & 6 & 14 \\ 0 & 4 & -9 & 13 \end{pmatrix}$

12. $\begin{pmatrix} 0 & 3 & -6 & 6 \\ 2 & -2 & 4 & -6 \\ 6 & 1 & 6 & -5 \\ 3 & 6 & -4 & 8 \end{pmatrix}$

13. $\begin{pmatrix} 2 & 4 & -2 & 6 \\ 2 & -4 & -2 & -10 \\ 3 & 8 & -3 & 5 \\ 4 & 3 & -3 & -4 \end{pmatrix}$

14. $\begin{pmatrix} 0 & 1 & 3 & 5 \\ -1 & 0 & 7 & 9 \\ -3 & -7 & 0 & 11 \\ -5 & -9 & -11 & 0 \end{pmatrix}$

15. $\begin{pmatrix} -3 & 5 & 5 & 5 \\ 5 & -3 & 5 & 5 \\ 5 & 5 & -3 & 5 \\ 5 & 5 & 5 & -3 \end{pmatrix}$

16. $\begin{pmatrix} 2 & 6 & 4 & 2 \\ 5 & 6 & 3 & 2 \\ 3 & -3 & 7 & -2 \\ 8 & 9 & 1 & -6 \end{pmatrix}$

17. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{pmatrix}$

18. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \\ 1 & 5 & 15 & 35 \end{pmatrix}$

19. $\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ -1 & 0 & 5 & 6 & 7 \\ -2 & -5 & 0 & 8 & 9 \\ -3 & -6 & -8 & 0 & 10 \\ -4 & -7 & -9 & -10 & 0 \end{pmatrix}$

20. $\begin{pmatrix} 2 & 4 & 1 & -4 & -2 \\ -1 & 1 & 0 & 8 & 2 \\ 3 & 8 & 0 & 2 & -5 \\ 4 & 9 & 0 & -3 & 1 \\ 5 & 13 & 0 & 7 & 6 \end{pmatrix}$

21. $\begin{pmatrix} 1 & 2 & -3 & -4 & 5 \\ 2 & 7 & -6 & 0 & 4 \\ -1 & -2 & 3 & 4 & -5 \\ 11 & 5 & 6 & 3 & 8 \\ -2 & 0 & 0 & 1 & -5 \end{pmatrix}$

22. $\begin{pmatrix} 8 & 2 & 0 & 6 & 4 \\ 4 & 5 & 7 & 3 & -2 \\ -4 & 0 & 8 & -3 & 1 \\ 0 & 9 & 5 & 0 & 4 \\ 12 & 2 & 3 & 9 & -6 \end{pmatrix}$

23. $\begin{pmatrix} -2 & 3 & 3 & 3 & 3 \\ 3 & -2 & 3 & 3 & 3 \\ 3 & 3 & -2 & 3 & 3 \\ 3 & 3 & 3 & -2 & 3 \\ 3 & 3 & 3 & 3 & -2 \end{pmatrix}$

In exercises 24 - 29 answer true or false and give an explanation.

24. If A is a 3×3 matrix and the vector $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is in $null(A)$ then the **determinant** of the **transpose** of A is zero, $det(A^{Tr}) = 0$.

25. If A is a 3×3 matrix and the vector $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is in $null(A)$ then the **determinant** of the **inverse** of A is zero, $det(A^{-1}) = 0$.

26. If the matrix A is an **invertible matrix** and $B = A^2$ then $det(B) > 0$.

27. For any two square matrices A and B, $det(AB) = det(BA)$.

28. If A is a **symmetric matrix** then $det(A) \geq 0$.

29. If A, B are **invertible** $n \times n$ matrices then $det(ABA^{-1}B^{-1}) = 1$.

Challenge Exercises (Problems)

1. Prove that the **determinant** of the $n \times n$ **identity matrix, I_n** is 1.

2. Prove that $det(D_i^n(c)) = c$. Here $D_i^n(c)$ is the scaling matrix obtained from the $n \times n$ **identity matrix** by multiplying the i^{th} row by c.

3. Prove **Theorem** (4.2.15). Specifically, the **determinant** of a **square matrix** A can be computed using a cofactor expansion in any row or column: If the $(i, j)-$entry of A is a_{ij} and the $(i, j)-$**cofactor** is C_{ij} then

$$det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

$$det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

Hints: Use **Theorem** (4.2.3) and **Theorem** (4.2.14).

4. Let J_n denote the $n \times n$ matrix all of whose entries are 1's. Let $a, b \in \mathbb{R}$. Prove that $det(aI_n + bJ_n) = a^{n-1}(a + bn)$.

5. Let n be odd and assume that A is a skew symmetric $n \times n$ matrix, that is, $A^T = -A$. Prove that $det(A) = 0$.

6. Let A be an $n \times n$ matrix and Q an invertible $n \times n$ matrix. Prove that

$det(QAQ^{-1}) = det(A)$.

7. Let A be an $m \times n$ matrix. Prove that the $col(A) = \mathbb{R}^m$ if and only if there is a choice of columns $\boldsymbol{v}_{i_1}, \boldsymbol{v}_{i_2}, \dots, \boldsymbol{v}_{i_m}$ such that the determinant of the matrix $(\boldsymbol{v}_{i_1} \boldsymbol{v}_{i_2} \dots \boldsymbol{v}_{i_m})$ is non-zero.

Quiz Solutions

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

See **the definition for the determinant of a 2 x 2 matrix**.

2. $det \begin{pmatrix} -10 & -15 \\ 3 & 4 \end{pmatrix} = 5.$

See **the definition for the determinant of a 2 x 2 matrix**.

3. $det \begin{pmatrix} 1 & 2 & 3 \\ 2 & 7 & 5 \\ 2 & 1 & 6 \end{pmatrix} = -3.$

See **the definition for the determinant of a 3x3 matrix**.

4. $det \begin{pmatrix} 2 & 7 & 5 \\ 1 & 2 & 3 \\ 2 & 1 & 6 \end{pmatrix} = 3.$

See **the definition for the determinant of a 3x3 matrix**.

5. $det \begin{pmatrix} 2 & 0 & 0 & 0 \\ -2 & 2 & 7 & 5 \\ 2 & 1 & 2 & 3 \\ -4 & 2 & 1 & 6 \end{pmatrix} = 6.$

See **recursive definition of the determinant**.

4.3. The Adjoint of a Matrix and Cramer's Rule

The adjoint of a **square matrix** is defined and Cramer's rule for solving a **linear system** of n equations in n variables is proved.

Am I Ready for This Material

The following are concepts used extensively in this section:

linear system

Solution of a linear system

invertible matrix

Inverse of an invertible matrix

Cofactors of a square matrix

Determinant of a square matrix

Quiz

Compute the determinants of the following matrices

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

2. $\begin{pmatrix} 2 & 5 & 8 \\ 3 & 5 & 7 \\ 4 & 5 & 6 \end{pmatrix}$

3. $\begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 5 & 2 & 5 \\ 3 & 5 & 8 & 5 \\ 2 & 3 & 2 & 6 \end{pmatrix}$

Quiz Solutions

New Concepts

In this section we introduce exactly one new concept: the **adjoint of a square matrix** A.

Theory (Why It Works)

Recall, in the previous section we proved **Theorem** (4.2.15) which states that the cofactor formula for the determinant works with any row or any column. More specifically, if A is an $n \times n$ matrix with entries a_{ij} and **cofactors** $C_{ij} = C_{ij}(A)$ then the **determinant** can be computed by expanding along the i^{th} row or the jth column:

$$det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

$$det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

Remark 4.9. When A is a 2×2 matrix the "submatrices" of A are just the entries of A. By convention we identify a scalar c with a 1×1 matrix. Moreover, we define the determinant of a 1×1 matrix (c) to be c.

Example 4.3.1. Consider the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ Let A' be the matrix obtained from A by replacing the third row by the second row so that $A' = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 4 & 5 & 6 \end{pmatrix}$.

Since two rows are identical, $det(A') = 0$ by **Theorem** (4.2.5). On the other hand let's see what we get if we compute the **determinant** of A' by **cofactor expansion along the third row**. Note that the **cofactors** for the third row of A' are the same as cofactors for the third row of A. However, instead of multiplying the cofactors by the corresponding entries of the third row of A we are multiplying by the corresponding entries of the second row A. Thus,

$$0 = det(A') = 4(12 - 15) - 5(6 - 12) + 6(5 - 8).$$

In terms of the **cofactors**

$$0 = det(A') = a_{21}C_{31} + a_{22}C_{32} + a_{23}C_{33}.$$

In exactly the same way for any $i \neq k$ we get

$$a_{i1}C_{k1} + a_{i2}C_{k2} + a_{i3}C_{k3} = 0.$$

If $j \neq l$ a similar formula holds for columns:

$$a_{1j}C_{1l} + a_{2j}C_{2l} + a_{3j}C_{3l} = 0.$$

There is nothing special about the case $n = 3$. These relations hold for any n. We therefore have the following result:

Theorem 4.3.1. *Let A be an $n \times n$ matrix with entries a_{ij} and **cofactors** C_{ij}. Then for $i \neq k$ and $j \neq l$ we have the following:*

$$a_{k1}C_{i1} + a_{k2}C_{i2} + \dots a_{kn}a_{in} = 0$$

$$a_{1l}C_{1j} + a_{2l}C_{2j} + \dots a_{nl}a_{nj} = 0.$$

Definition 4.8. The ***adjoint*** of the **square matrix** A with entries a_{ij} and **cofactors** C_{ij} is the matrix whose $(i, j)-$entry is C_{ji}.

In other words, the adjoint is the **transpose** of the matrix whose $(i, j)-$entry is the **cofactor** C_{ij}.

Example 4.3.2. Let $A = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 7 & 6 \\ 4 & 9 & 8 \end{pmatrix}$.

We compute the cofactors:

$$C_{11} = (-1)^{1+1}det\begin{pmatrix} 7 & 6 \\ 9 & 8 \end{pmatrix} = 2, C_{12} = (-1)^{1+2}det\begin{pmatrix} 2 & 6 \\ 4 & 8 \end{pmatrix} = 8$$

$$C_{13} = (-1)^{1+3} det \begin{pmatrix} 2 & 7 \\ 4 & 9 \end{pmatrix} = -10, C_{21} = (-1)^{2+1} det \begin{pmatrix} 3 & 5 \\ 9 & 8 \end{pmatrix} = 21$$

$$C_{22} = (-1)^{2+2} det \begin{pmatrix} 1 & 5 \\ 4 & 8 \end{pmatrix} = -12, C_{23} = (-1)^{2+3} det \begin{pmatrix} 1 & 3 \\ 4 & 9 \end{pmatrix} = 3$$

$$C_{31} = (-1)^{3+1} det \begin{pmatrix} 3 & 5 \\ 7 & 6 \end{pmatrix} = -17, C_{32} = (-1)^{3+2} det \begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix} = 4$$

$$C_{33} = (-1)^{3+3} det \begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix} = 1.$$

Therefore $Adj(A) = \begin{pmatrix} 2 & 21 & -17 \\ 8 & -12 & 4 \\ -10 & 3 & 1 \end{pmatrix}$.

We compute the products of A and its **adjoint**.

$$Adj(A)A = \begin{pmatrix} 2 & 21 & -17 \\ 8 & -12 & 4 \\ -10 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 \\ 2 & 7 & 6 \\ 4 & 9 & 8 \end{pmatrix} = \begin{pmatrix} -24 & 0 & 0 \\ 0 & -24 & 0 \\ 0 & 0 & -24 \end{pmatrix} =$$

$$-24 \times I_3 = det(A) \times I_3.$$

In exactly the same fashion we can show that this holds for multiplication on the right:

$$AAdj(A) = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 7 & 6 \\ 4 & 9 & 8 \end{pmatrix} \begin{pmatrix} 2 & 21 & -17 \\ 8 & -12 & 4 \\ -10 & 3 & 1 \end{pmatrix} = \begin{pmatrix} -24 & 0 & 0 \\ 0 & -24 & 0 \\ 0 & 0 & -24 \end{pmatrix} =$$

$$-24 \times I_3 = det(A) \times I_3.$$

There is nothing special about this particular matrix or 3×3 matrices for that matter. In fact, this relationship holds in general:

Theorem 4.3.2. *Let A be an $n \times n$ matrix with entries a_{ij} and* ***cofactors*** *C_{ij}. Then $Adj(A)A = det(A)I_n = AAdj(A)$.*

Proof. The i^{th} diagonal entry of $Adj(A)A$ is the product of the i^{th} row of $Adj(A)$ and the i^{th} column of A and this is $a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$ which is equal to $det(A)$ by **Theorem** (4.2.15).

On the other hand, for $i \neq j$ the $(i, j)-$entry of $Adj(A)A$ is the of the i^{th} row of $Adj(A)$ and the j^{th} column of A. This gives $a_{1j}C_{1i} + a_{2j}C_{2i} + \ldots a_{nj}C_{ni}$ which is equal to zero by **Theorem** (4.3.1). Thus, all the entries of $Adj(A)A$ off the **main diagonal** is zero and each **diagonal entry** is $det(A)$. Thus, $Adj(A)A = det(A)I_n$ as claimed. That $AAdj(A) = det(A)I_n$ is proved similarly. $\square$

An immediate consequence of **Theorem** (4.3.2) is the following:

Theorem 4.3.3. *Assume A is an $n \times n$ **invertible matrix**. Then $A^{-1} = \frac{1}{det(A)}Adj(A)$.*

Example 4.3.3. The **inverse** of the matrix A of **Example** (4.3.2) is

$$\begin{pmatrix} -\frac{1}{12} & -\frac{7}{8} & \frac{17}{24} \\ -\frac{1}{3} & \frac{1}{2} & -\frac{1}{6} \\ \frac{5}{12} & -\frac{1}{8} & -\frac{1}{24} \end{pmatrix}$$

This expression obtained for the **inverse** of a matrix A from the **adjoint** of A can be used get a formula for the solution of a **linear system** of n equations in n variables when the **coefficient matrix** is A. This result goes by the name of "**Cramer's Rule**".

Theorem 4.3.4. *Let A be an $n \times n$ **invertible matrix** and $\boldsymbol{b}$ vector in $\mathbb{R}^n$. Let A_j denote the matrix obtained from A by replacing the j^{th} column of A with the $n-$vector $\boldsymbol{b}$. Set $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Then the **linear system** represented by the matrix equation*

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

*has a unique **solution** given by*

$$x_1 = \frac{det(A_1)}{det(A)}, x_2 = \frac{det(A_2)}{det(A)}, \ldots, x_n = \frac{det(A_n)}{det(A)}.$$

Proof.

$$\boldsymbol{x} = A^{-1}\boldsymbol{b} = \frac{1}{det(A)}\begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \dots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Then x_j is the product of the j^{th} row of A^{-1} with $\boldsymbol{b}$ and so

$$x_j = b_1\frac{C_{1j}}{det(A)} + b_2\frac{C_{2j}}{det(A)} + \dots + b_n\frac{C_{nj}}{det(A)} =$$

$$\frac{1}{det(A)}[b_1C_{1j} + b_2C_{2j} + \dots + b_nC_{nj}] \tag{4.9}$$

Now suppose that A_j is the matrix obtained from A by replacing the j^{th} column of A by $\boldsymbol{b}$ so that

$$A_j = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,j-1} & b_1 & a_{1,j+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2,j-1} & b_2 & a_{2,j+1} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{n,j-1} & b_n & a_{n,j+1} & \dots & a_{nn} \end{pmatrix}$$

If we compute the **determinant** of A_j using the **cofactor expansion in the j^{th} column** we obtain

$$det(A_j) = b_1C_{1j}(A_j) + b_2C_{2j}(A_j) + \dots + b_nC_{nj}(A_j)$$

where $C_{ij}(A_j)$ is the **(i,j)-cofactor** of A_j. However, since A_j and A are the same matrix except for the j^{th} column, $C_{ij}(A_j) = C_{ij}$. Therefore

$$det(A_j) = b_1C_{1j} + b_2C_{2j} + \dots + b_nC_{nj} \tag{4.10}$$

It now follows from **Equation** (4.9) and **Equation** (4.10) that $x_j = \frac{det(A_i)}{det(A)}$. □

Example 4.3.4. Consider the **linear system** represented by the matrix equation

$$A\boldsymbol{x} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 7 & 6 \\ 4 & 9 & 8 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 11 \\ 19 \end{pmatrix}.$$

We have seen that the **determinant** of A is -24. Then **Cramers rule** states the unique **solution** to this **linear system** is:

$$x_1 = \frac{det\begin{pmatrix} 10 & 3 & 5 \\ 11 & 7 & 6 \\ 19 & 9 & 8 \end{pmatrix}}{-24}$$

$$x_2 = \frac{det\begin{pmatrix} 1 & 10 & 5 \\ 2 & 11 & 6 \\ 5 & 10 & 8 \end{pmatrix}}{-24}$$

$$x_3 = \frac{det\begin{pmatrix} 1 & 3 & 10 \\ 2 & 7 & 11 \\ 4 & 9 & 19 \end{pmatrix}}{-24}.$$

This yields

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{-72}{-24} \\ \frac{24}{-24} \\ \frac{-48}{-24} \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$$

as can be checked.

What You Can Now Do

1. Given an $n \times n$ matrix compute the **adjoint matrix**, $Adj(A)$, of A.

2. When A is an $n \times n$ **invertible matrix** compute the **inverse** of A making use of the **adjoint matrix** of A.

3. When A is an $n \times n$ **invertible matrix** find the unique **solution** to the **linear system** represented by the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ using **Cramer's Rule**.

Method (How To Do It)

Method 4.3.1. Given an $n \times n$ matrix compute the **adjoint matrix**, $Adj(A)$, of A.

If the **cofactors** of A are $C_{ij} = C_{ij}(A)$ then the adjoint matrix is

$$\begin{pmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \dots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{pmatrix}$$

Example 4.3.5. Let $A = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 7 & 8 \\ 1 & 1 & 7 \end{pmatrix}$. The **cofactors** of A are:

$$C_{11} = det \begin{pmatrix} 7 & 8 \\ 1 & 7 \end{pmatrix} = 41, C_{12} = -det \begin{pmatrix} 2 & 8 \\ 1 & 7 \end{pmatrix} = -6, C_{13} = det \begin{pmatrix} 2 & 7 \\ 1 & 1 \end{pmatrix} = -5,$$

$$C_{21} = -det \begin{pmatrix} 3 & 5 \\ 1 & 7 \end{pmatrix} = -16, C_{22} = det \begin{pmatrix} 1 & 5 \\ 1 & 7 \end{pmatrix} = 2, C_{23} = -det \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix} = 2,$$

$$C_{31} = det \begin{pmatrix} 3 & 5 \\ 7 & 8 \end{pmatrix} = -11, C_{32} = -det \begin{pmatrix} 1 & 5 \\ 2 & 8 \end{pmatrix} = 2, C_{33} = det \begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix} = 1.$$

$$Adj(A) = \begin{pmatrix} 41 & -16 & -11 \\ -6 & 2 & 2 \\ -5 & 2 & 1 \end{pmatrix}$$

Example 4.3.6. Let $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 3 & 3 \\ 4 & 3 & 4 & 3 \\ 3 & 2 & 2 & 6 \end{pmatrix}$. Then the **cofactors** are:

$$C_{11} = det \begin{pmatrix} 3 & 3 & 3 \\ 3 & 4 & 3 \\ 2 & 2 & 6 \end{pmatrix} = 12, C_{12} = -det \begin{pmatrix} 2 & 3 & 3 \\ 4 & 4 & 3 \\ 3 & 2 & 6 \end{pmatrix} = 21,$$

$$C_{13} = det \begin{pmatrix} 2 & 3 & 3 \\ 4 & 2 & 3 \\ 3 & 2 & 6 \end{pmatrix} = -24, C_{14} = -det \begin{pmatrix} 2 & 3 & 3 \\ 4 & 3 & 4 \\ 3 & 2 & 2 \end{pmatrix} = -5,$$

$$C_{21} = -det \begin{pmatrix} 1 & 1 & 1 \\ 3 & 4 & 3 \\ 2 & 2 & 6 \end{pmatrix} = -4, C_{22} = det \begin{pmatrix} 1 & 1 & 1 \\ 4 & 4 & 3 \\ 3 & 2 & 6 \end{pmatrix} = -1,$$

$$C_{23} = -det \begin{pmatrix} 1 & 1 & 1 \\ 4 & 3 & 3 \\ 3 & 2 & 6 \end{pmatrix} = 4, C_{24} = det \begin{pmatrix} 1 & 1 & 1 \\ 4 & 3 & 4 \\ 3 & 2 & 2 \end{pmatrix} = 1,$$

$$C_{31} = det\begin{pmatrix}1 & 1 & 1\\ 3 & 3 & 3\\ 2 & 2 & 6\end{pmatrix} = 0, C_{32} = -det\begin{pmatrix}1 & 1 & 1\\ 2 & 3 & 3\\ 3 & 2 & 6\end{pmatrix} = -4,$$

$$C_{33} = det\begin{pmatrix}1 & 1 & 1\\ 2 & 3 & 3\\ 3 & 2 & 6\end{pmatrix} = 4, c_{34} = -det\begin{pmatrix}1 & 1 & 1\\ 2 & 3 & 3\\ 3 & 2 & 2\end{pmatrix} = 0,$$

$$C_{41} = -det\begin{pmatrix}1 & 1 & 1\\ 3 & 3 & 3\\ 3 & 4 & 3\end{pmatrix} = 0, C_{42} = det\begin{pmatrix}1 & 1 & 1\\ 2 & 3 & 3\\ 4 & 4 & 3\end{pmatrix} = -1,$$

$$C_{43} = -det\begin{pmatrix}1 & 1 & 1\\ 2 & 3 & 3\\ 4 & 3 & 3\end{pmatrix} = 0, C_{44} = det\begin{pmatrix}1 & 1 & 1\\ 2 & 3 & 3\\ 4 & 3 & 4\end{pmatrix} = 1.$$

$$Adj(A) = \begin{pmatrix}12 & -4 & 0 & 0\\ 21 & -1 & -4 & -1\\ -24 & 4 & 4 & 0\\ -5 & 1 & 0 & 1\end{pmatrix}.$$

Example 4.3.7. Let $A = \begin{pmatrix}1 & 2 & 1 & 2\\ 2 & 5 & 2 & 5\\ 1 & 2 & 1 & 2\\ 2 & 3 & 2 & 3\end{pmatrix}$.

Eight cofactors $C_{21}, C_{22}, C_{23}, C_{24}, C_{41}, C_{42}, C_{43}, C_{44}$ are visibly seen to be zero since two rows will be equal. Also, $C_{11}, C_{12}, C_{13}, C_{14}, C_{31}, C_{32}, C_{33}, C_{34}$ are zero since these submatrices each has two equal columns. Therefore the $Adj(A) = \mathbf{0}_{4\times 4}$.

Method 4.3.2. When A is an $n \times n$ **invertible matrix** compute the **inverse** of A making use of the **adjoint matrix** of A.

Compute the **determinant** of A to determine if A is an **invertible matrix**. If $det(A) = 0$, stop, A is **non-nvertible**.

If $det(A) \neq 0$ then A is **invertible**. Compute the **adjoint matrix** of A as in **Method** (4.3.1) and then divide all the entries of $Adj(A)$ by $det(A)$.

Example 4.3.8. Determine if $A = \begin{pmatrix}1 & 2 & 3\\ 2 & 5 & 7\\ 3 & 5 & 9\end{pmatrix}$ is **invertible** by calculating the **determinant**. If A is **invertible** use the **adjoint** of A to compute the **inverse** of A.

$$det(A) = det(\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 5 & 9 \end{pmatrix}) = det(\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & -1 & 0 \end{pmatrix}) = det(\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}) = 1$$

$$C_{11} = det(\begin{pmatrix} 5 & 7 \\ 5 & 9 \end{pmatrix}) = 10, C_{12} = -det(\begin{pmatrix} 2 & 7 \\ 3 & 9 \end{pmatrix}) = 3, C_{13} = det(\begin{pmatrix} 2 & 5 \\ 3 & 5 \end{pmatrix}) = -5,$$

$$C_{21} = -det(\begin{pmatrix} 2 & 3 \\ 5 & 9 \end{pmatrix}) = -3, C_{22} = det(\begin{pmatrix} 1 & 3 \\ 3 & 9 \end{pmatrix}) = 0, C_{23} = -det(\begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}) = 1,$$

$$C_{31} = det(\begin{pmatrix} 2 & 3 \\ 5 & 7 \end{pmatrix}) = -1, C_{32} = -det(\begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix}) = -1, C_{33} = det(\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}) = 1.$$

Since $det(A) = 1, A^{-1} = Adj(A) = \begin{pmatrix} 10 & -3 & -1 \\ 3 & 0 & -1 \\ -5 & 1 & 1 \end{pmatrix}$

Example 4.3.9. Determine if $A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 4 & 3 \\ 3 & 1 & 4 \end{pmatrix}$ is **invertible** by calculating the **determinant**. If A is **invertible** use the **adjoint** of A to compute the **inverse** of A.

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

$$(-2)(-1)det(\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 2 & -1 \end{pmatrix}) = (-2)(-1)det(\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -3 \end{pmatrix}) = (-2)(-1)(-3) = -6.$$

A is **invertible** since the **determinant** is $-6 \neq 0$.

$$Adj(A) = \begin{pmatrix} 13 & -2 & -5 \\ 1 & -2 & 1 \\ -10 & 2 & 2 \end{pmatrix}$$

$$A^{-1} = \frac{1}{det(A)} Adj(A) = -\frac{1}{6} \begin{pmatrix} 13 & -2 & -5 \\ 1 & -2 & 1 \\ -10 & 2 & 2 \end{pmatrix} = \begin{pmatrix} -\frac{13}{6} & \frac{1}{3} & \frac{5}{6} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} \\ \frac{5}{3} & -\frac{1}{3} & -\frac{1}{3} \end{pmatrix}$$

Method 4.3.3. When A is an $n \times n$ **invertible matrix** find the unique **solution** to the **linear system** represented by the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ using **Cramer's Rule**.

The unique **solution** is the vector $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ with $x_j = \frac{det(A_j)}{det(A)}$ where A_j is the matrix obtained from A by replacing the j^{th} column by $\boldsymbol{b}$.

Example 4.3.10. Use **Cramer's rule** to solve the following **linear system**

$$\begin{array}{ccccccc} x & + & y & + & 2z & = & 2 \\ 2x & + & 3y & + & 5z & = & 3 \\ 3x & + & 2y & + & 3z & = & 3 \end{array}$$

The **coeffient matrix** of the **linear system** is $\begin{pmatrix} 1 & 1 & 2 \\ 2 & 3 & 5 \\ 3 & 2 & 3 \end{pmatrix}$. We compute its **determinant** :

$$det \begin{pmatrix} 1 & 1 & 2 \\ 2 & 3 & 5 \\ 3 & 2 & 3 \end{pmatrix} = det \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & -1 & -3 \end{pmatrix} = det \begin{pmatrix} 1 & 1 \\ -1 & -3 \end{pmatrix} = -3 + 1 = -2.$$

The solution, by **Cramer's rule**, is

$$x_1 = \frac{det \begin{pmatrix} 2 & 1 & 2 \\ 3 & 3 & 5 \\ 3 & 2 & 3 \end{pmatrix}}{-2} = \frac{-2}{-2} = 1$$

$$x_2 = \frac{det \begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 5 \\ 3 & 3 & 3 \end{pmatrix}}{-2} = \frac{6}{-2} = -3$$

$$x_3 = \frac{det \begin{pmatrix} 1 & 1 & 2 \\ 2 & 3 & 3 \\ 3 & 2 & 3 \end{pmatrix}}{-2} = \frac{-4}{-2} = 2.$$

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

Exercises In 1-6 compute the **adjoint** of the given matrix. See **Method** (4.3.1).

1. $\begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}$

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

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

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

5. $\begin{pmatrix} 1 & -2 & 5 \\ -1 & 2 & -5 \\ 2 & -4 & 10 \end{pmatrix}$

6. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 4 & 6 & 8 \end{pmatrix}$

In 7-10 determine if the matrix A is **invertible** and, if so, use the **adjoint matrix** of A to compute the **inverse**. See **Method** (4.3.2).

7. $\begin{pmatrix} 2 & 5 \\ 3 & 7 \end{pmatrix}$

8. $\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}$

9. $\begin{pmatrix} 1 & 3 & -3 \\ 2 & 7 & -7 \\ 1 & 1 & -2 \end{pmatrix}$

10. $\begin{pmatrix} 1 & 2 & -3 \\ 3 & 4 & -7 \\ 2 & 6 & -7 \end{pmatrix}$

In 11-13 solve the **linear system** using **Cramer's rule**. See **Method** (4.3.3).

11.

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

12.

$$\begin{array}{ccccccc} 2x & - & y & + & 3z & = & 3 \\ 3x & - & 2y & + & 2z & = & 6 \\ 5x & - & 2y & + & 9z & = & 7 \end{array}$$

13.

$$\begin{array}{ccccccc} x & + & y & - & 2z & = & 0 \\ 2x & + & 3y & - & 3z & = & -1 \\ 2x & + & y & - & 3z & = & 3 \end{array}$$

Challenge Exercises (Problems)

1. Assume that A is an $n \times n$ **non-invertible matrix**. Prove that $det(Adj(A)) = 0$.

2. If A is an $n \times n$ matrix and $det(A) = d$ prove that $det(Adj(A)) = d^{n-1}$.

Quiz Solutions

1. $det \begin{pmatrix} 2 & 2 & 4 \\ 3 & 6 & 3 \\ 2 & 3 & 4 \end{pmatrix} = (2)(3)det \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 3 & 4 \end{pmatrix} =$

$(2)(3)det \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \end{pmatrix} = (2)(3)det \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 11 \end{pmatrix} = 6.$

Not right, see **Method** (4.2.1).

2. $det \begin{pmatrix} 2 & 5 & 8 \\ 3 & 5 & 7 \\ 4 & 5 & 6 \end{pmatrix} = (5)det \begin{pmatrix} 2 & 1 & 8 \\ 3 & 1 & 7 \\ 4 & 1 & 6 \end{pmatrix} =$

$$(5)det(\begin{pmatrix} 2 & 1 & 8 \\ 1 & 0 & -1 \\ 0 & -1 & -10 \end{pmatrix}) = (5)(-1)det\begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 8 \\ 0 & -1 & -10 \end{pmatrix} =$$

$$(5)(-1)det\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 10 \\ 0 & -1 & -10 \end{pmatrix} = (5)(-1)det\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 10 \\ 0 & 0 & 0 \end{pmatrix} = 0.$$

Not right, see **Method** (4.2.1) and **Corollary** (4.2.6).

3. $det\begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 5 & 2 & 5 \\ 3 & 5 & 8 & 5 \\ 2 & 3 & 2 & 6 \end{pmatrix} = det\begin{pmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 5 & -1 \\ 0 & -1 & 0 & 2 \end{pmatrix}) = det\begin{pmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} = 15.$

Not right, see **Method** (4.2.1)

Chapter 5

Abstract Vector Spaces

5.1. Introduction to Abstract Vector Spaces

In this section we define the notion of a vector space and give multiple examples.

Am I Ready for This Material

Extensive use is made of the following concepts and procedures in this section and they need to be mastered to fully understand the new ideas and methods introduced here:

linear system

augmented matrix of a linear system

consistent linear system

reduced echelon form of a matrix

$\mathbb{R}^n$

linear combination of a sequence of vectors

subspace of $\mathbb{R}^n$

equality of matrices

Quiz

In 1 - 2 let $m(a,b,c) = \begin{pmatrix} 2a-3b+c & 2a-4b+2c \\ a+b+c & 3a-2b+2c \end{pmatrix}$.

1. Determine if there is a choice of (a,b,c) so that $m(a,b,c) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

2. Determine if there is a choice of (a,b,c) so that $m(a,b,c) = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$.

3. Let $m(a,b,c) = \begin{pmatrix} a+2b+c+1 & 2a+b+3c \\ a+2b+2c & 2a+b+3c \end{pmatrix}$. Determine if there is a choice of (a,b,c) such that $m(a,b,c) = \overline{0}_{2\times 2}$.

Quiz Solutions

New Concepts

In much of what follows we will be studying abstract vector spaces of which $\mathbb{R}^n$ is an example. Also arising in this section are several related definitions.

vector space

linear combination of a sequence of vectors in a vector space

the vector space $M_{m\times n}$ of $m \times n$ matrices

the vector space $\mathbb{R}[x]$ of polynomials in the indeterminate x

the vector space $F(\mathbb{R})$

subspace of a vector space

the zero subspace of a vector space

sum of subspaces of a vector space

Theory (Why It Works)

We have seen the following definition in some form several times. Now we begin to study it more seriously. Basically, it is abstracted from properties of $\mathbb{R}^n$.

Definition 5.1. Let V be a set equipped with two operations: ***addition*** and ***scalar multiplication***. Addition is a rule (function) that associates to any pair $(\boldsymbol{u}, \boldsymbol{v})$ of elements from V a third element of V called the ***sum*** of $\boldsymbol{u}$ and $\boldsymbol{v}$ and denoted by $\boldsymbol{u} + \boldsymbol{v}$.

Scalar multiplication is a rule (function) which associates to each pair $(c, \boldsymbol{u})$ consisting of a scalar c and element $\boldsymbol{u}$ of V an element in V called the ***scalar multiple*** of $\boldsymbol{u}$ by c and denoted by $c\boldsymbol{u}$. V is said to be a ***vector space*** if these operations satisfy the following axioms:

(A1) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ for every $\boldsymbol{u}, \boldsymbol{v} \in V$. **Addition is commutative**.

(A2) $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$ for every $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in V. **Addition is associative**.

(A3) There is a special element $\mathbf{0}_V$ called the **zero vector** such that $\boldsymbol{u} + \mathbf{0}_V = \boldsymbol{u}$ for every $\boldsymbol{u} \in V$. This is the **existence of additive identity**.

(A4) Every element $\boldsymbol{u}$ in V has a ***negative***, denoted by $-\boldsymbol{u}$ such that $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}_V$. This is the **existence of additive inverses**.

(M1) For every scalar a and pair $\boldsymbol{u}, \boldsymbol{v}$ from V, $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$. This is a form of the **distributive property**.

(M2) For every pair of scalars a, b and element $\boldsymbol{u}$ of V, $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$. This is another form of the **distributive property**.

(M3) $(ab)\boldsymbol{u} = a(b\boldsymbol{u})$ for all scalars a and b and vectors $\boldsymbol{u}$.

(M4) $1\boldsymbol{u} = \boldsymbol{u}$.

Elements of V are referred to as ***vectors***. The axioms (M1) and (M2) are the ***distributive laws***. Finally, note that the definition hypothesizes the existence of a certain element, the zero vector as in (A3) and therefore, V is not the empty set (which otherwise satisfies all the axioms).

Remark 5.1. Some books make it a separate assumption that $\boldsymbol{u} + \boldsymbol{v}$ is in V whereas this is explicit in our definition of an addition. Likewise these same books would have an axiom to state that the scalar product, $c\boldsymbol{u}$ is in V while we take this to be part of a meaning of a scalar product.

The following should be familar since it is exactly as we defined it for $\mathbb{R}^n$.

Definition 5.2. Let V be a **vector space** and $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ a sequence of vectors from V. A ***linear combination*** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is an expression of the form $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots c_k\boldsymbol{v}_k$ where $c_1, \ldots, c_k$ are scalars.

In a moment we will prove some abstract results, however for the time being let us give some examples.

Example 5.1.1. $\mathbb{R}^n$ is a vector space.

Definition 5.3. We denote by $M_{m\times n}(\mathbb{R})$ the collection of all $m \times n$ matrices with entries taken from the real numbers, $\mathbb{R}$.

Example 5.1.2. $M_{m\times n}(\mathbb{R})$ with **addition** and **scalar multiplication** defined in Section (3.3) is a **vector space**.

Definition 5.4. We denote by $\mathbb{R}[x]$ the collection of all polynomials in the indeterminate x with real coefficients.

Example 5.1.3. $\mathbb{R}[x]$ with the usual addition of polynomials and multiplication by constants is a **vector space**.

Definition 5.5. Denote by $F(\mathbb{R})$ the collection of all functions from $\mathbb{R}$ to $\mathbb{R}$. $F(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R}\}$.

Example 5.1.4. For two function f, g in $F(\mathbb{R})$ define addition by $(f+g)(x) = f(x)+g(x)$, that is, the pointwise addition of functions. Likewise scalar multiplication is given by $(cf)(x) = cf(x)$. With these operations $F(\mathbb{R})$ becomes is a **vector space**. The additive identity $\mathbf{0}_{F(\mathbb{R})}$ is the constant function $\mathbf{0}$ which takes every element of $\mathbb{R}$ to zero. The negative of a function $f : \mathbb{R} \to \mathbb{R}$ is the function $-f = (-1)f$ which satisfies, $(-f)(x) = -f(x)$.

Example 5.1.5. All solutions of the differential equation $\frac{d^2y}{dx^2}+y = 0$. Since solutions to the equation are functions we use the addition and scalar multiplication introduced in **Example** (5.1.4). Note that there are solutions since, in particular, $sin\ x, cos\ x$ satisfy this differential equation.

We now come to some basic results.

Theorem 5.1.1. *Let V be a* **vector space**.

1. The element $\mathbf{0}_V$ in V is unique. By this we mean if an element $\boldsymbol{e}$ of V satisfies $\boldsymbol{u}+\boldsymbol{e} = \boldsymbol{e}+\boldsymbol{u} = \boldsymbol{u}$ for every vector $\boldsymbol{u}$ in V then $\boldsymbol{e} = \mathbf{0}_V$.

2. The **negative** *of a vector $\boldsymbol{u}$ is unique, that is, if $\boldsymbol{v}$ is a vector which satisifes $\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{v}+\boldsymbol{u} = \mathbf{0}_V$ then $\boldsymbol{v} = -\boldsymbol{u}$.*

Proof. 1. Suppose that

$$\boldsymbol{u}+\boldsymbol{e} = \boldsymbol{e}+\boldsymbol{u} = \boldsymbol{u} \tag{5.1}$$

for every $\boldsymbol{u}$ in V. We already know that

$$\boldsymbol{u}+\mathbf{0}_V = \mathbf{0}_v+\boldsymbol{u} = \boldsymbol{u} \tag{5.2}$$

for every vector $\boldsymbol{u}$ in V.

Consider the vector $\mathbf{0}_V+\boldsymbol{e}$. By (5.2) we have $\mathbf{0}_V+\boldsymbol{e} = \mathbf{0}_V$ and by (5.1) that $\mathbf{0}_V+\boldsymbol{e} = \boldsymbol{e}$. Thus, $\boldsymbol{e} = \mathbf{0}_V$.

2. Suppose

$$\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{v}+\boldsymbol{u} = \mathbf{0}_V. \tag{5.3}$$

We know that

$$\boldsymbol{u}+(-\boldsymbol{u}) = (-\boldsymbol{u})+\boldsymbol{u} = \mathbf{0}_V. \tag{5.4}$$

Consider the sum $(-\boldsymbol{u})+(\boldsymbol{u}+\boldsymbol{v})$. By (5.3) we have $(-\boldsymbol{u})+(\boldsymbol{u}+\boldsymbol{v}) = (-\boldsymbol{u})+\mathbf{0}_V = -\boldsymbol{u}$. However, by **associativity of addition** we have

$$(-\boldsymbol{u}) + (\boldsymbol{u} + \boldsymbol{v}) = [(-\boldsymbol{u}) + \boldsymbol{u}] + \boldsymbol{v}$$

which by (5.4) is equalt to $\mathbf{0}_V + \boldsymbol{v} = \boldsymbol{v}$. Therefore $-\boldsymbol{u} = \boldsymbol{v}$. □

Our next result makes explicit several basic properties that hold in any **vector space**.

Theorem 5.1.2. *Let V be a **vector space**, $\boldsymbol{u}$ a vector in V, and c a scalar. Then the following hold:*

1. $0\boldsymbol{u} = \mathbf{0}_V$.

2. $c\mathbf{0}_V = \mathbf{0}_V$.

3. If $c\boldsymbol{u} = \mathbf{0}_V$ then either $c = 0$ or $\boldsymbol{u} = \mathbf{0}_V$.

4. $(-c)\boldsymbol{u} = -(c\boldsymbol{u})$.

Proof. 1. Let $\boldsymbol{v}$ be the negative of $0\boldsymbol{u}$ so that $\boldsymbol{v} + (0\boldsymbol{u}) = (0\boldsymbol{u}) + \boldsymbol{v} = \mathbf{0}_V$.

Since $0 = 0 + 0$ when we multiply by $\boldsymbol{u}$ we get

$$0\boldsymbol{u} = (0+0)\boldsymbol{u} = 0\boldsymbol{u} + 0\boldsymbol{u}. \tag{5.5}$$

Now add $\boldsymbol{v}$ to both sides of (5.5). This yields:

$$\boldsymbol{v} + (0\boldsymbol{u}) = \boldsymbol{v} + [0\boldsymbol{u} + 0\boldsymbol{u}] = [\boldsymbol{v} + (0\boldsymbol{u})] + 0\boldsymbol{u} \tag{5.6}$$

(the last step by **associativity**). On left side of (5.6) we have $\mathbf{0}_V$ and on the right hand side we have $\mathbf{0}_V + (0\boldsymbol{u}) = 0\boldsymbol{u}$. Thus, $0\boldsymbol{u} = \mathbf{0}_V$ as asserted.

2. This is left as a **challenge exercise**.

3. Assume

$$c\boldsymbol{u} = \mathbf{0}_V. \tag{5.7}$$

We assume that $c \neq 0$ and will show that $\boldsymbol{u} = \mathbf{0}_V$. Since $c \neq 0$ its multiplicative inverse (reciprocal) $\frac{1}{c}$ exists. We multiply both sides of (5.7) by $\frac{1}{c}$ to obtain

$$\frac{1}{c}[c\boldsymbol{u}] = \frac{1}{c}\mathbf{0}_V. \tag{5.8}$$

After applying **axiom (M3)** on the left hand side of (5.8) we get $[\frac{1}{c} \times c]\boldsymbol{u} = 1\boldsymbol{u} = \boldsymbol{u}$ by **axiom (M4)**. On the other hand, by part 2 of this theorem we know that $\frac{1}{c}\mathbf{0}_V = \mathbf{0}_V$ and therefore $\boldsymbol{u} = \mathbf{0}_V$ as claimed.

4. From 1) we know that

$$\mathbf{0}_V = 0\boldsymbol{u} \tag{5.9}$$

On the other hand, $c+(-c) = 0$. Substituting this on the right hand side of (5.9) yields

$$\mathbf{0}_V = [c + (-c)]\boldsymbol{u} \tag{5.10}$$

After **distributing** on the right hand side of (5.10) we get

$$\mathbf{0}_V = c\boldsymbol{u} + (-c)\boldsymbol{u}. \tag{5.11}$$

Now set $\boldsymbol{v} = -[c\boldsymbol{u}]$ so that $\boldsymbol{v} + c\boldsymbol{u} = c\boldsymbol{u} + \boldsymbol{v} = \mathbf{0}_V$ and add $\boldsymbol{v}$ to both sides of (5.11):

$$\boldsymbol{v} + \mathbf{0}_V = \boldsymbol{v} + [c\boldsymbol{u} + (-c)\boldsymbol{u}]. \tag{5.12}$$

On the left hand side of (5.12) we have $\boldsymbol{v} = -(c\boldsymbol{u})$. On the right hand side, after applying **associativity** we have $[\boldsymbol{v} + c\boldsymbol{u}] + [(-c)\boldsymbol{u}] = \mathbf{0}_V + (-c)\boldsymbol{u} = (-c)\boldsymbol{u}$. Thus, $-(c\boldsymbol{u}) = \boldsymbol{v} = (-c)\boldsymbol{u}$. □

Just as with $\mathbb{R}^n$, we will be interested in subsets of a **vector space** V which are themselves vector spaces (with the operations of addition and scalar multiplication "inherited" from V). This is the subject of the next defintition.

Definition 5.6. A subset W of a **vector space** V is called a ***subspace of*** V if W is itself a vector space under the addition and scalar multiplication inherited from V.

The next result gives simple criteria for a subset to be a **subspace**.

Theorem 5.1.3. *Let V be a* **vector space** *and W a non-empty subset of V. Then W is a* **subspace** *of V if and only if the following hold:*

1. For all $\boldsymbol{u}, \boldsymbol{v} \in W$ the **sum** *$\boldsymbol{u}+\boldsymbol{v}$ is in W. We say that W is* **closed under addition***.*

2. For every vector $\boldsymbol{u}$ in W and scalar c the **scalar product** *$c\boldsymbol{u}$ is in W. We say that W is* **closed under scalar multiplication***.*

Proof. Assume that W is a **subspace**. Let us denote the addition in W by $+_W$. By the definition of an addition in a vector space for $\boldsymbol{u}, \boldsymbol{v} \in W$, $\boldsymbol{u} +_W \boldsymbol{w}$ is an element in W. However, for two elements $\boldsymbol{u}, \boldsymbol{v}$ in W, $\boldsymbol{u} +_W \boldsymbol{v}$ is just $\boldsymbol{u} + \boldsymbol{v}$ (the addition in V.) It therefore follows for $\boldsymbol{u}, \boldsymbol{v} \in W$ that $\boldsymbol{u}+\boldsymbol{v} \in W$. In a similar fashion, for $\boldsymbol{u}$ in W and scalar c, $c\boldsymbol{u} \in W$. Thus, if W is a **subspace** of V then 1) and 2) hold.

Conversely, assume that W is non-empty (it has vectors) and that 1) and 2) hold. The **axioms (A1) and (A2)** hold since they hold in V and the addition in W is the same as the addition in V. For **axiom (A3)** we have to show that the **zero vector** of V belongs to W. We do know that W is non-empty so let $\boldsymbol{u} \in W$. By 2) we know for any scalar c that also $c\boldsymbol{u} \in W$. In particular, $0\boldsymbol{u} \in W$. However, by part 1) of **Theorem** (5.1.2), $0\boldsymbol{u} = \mathbf{0}_V$ and $\mathbf{0}_V \in W$. Since for all $\boldsymbol{v} \in V, \mathbf{0}_V + \boldsymbol{v} = \boldsymbol{v}$ it follows that this holds in W as well.

We also have to show that for any vector $\boldsymbol{u} \in W$ that the **negative** of $\boldsymbol{u}$ belongs to V. However, by 2) we know that $(-1)\boldsymbol{u} \in W$. By part 4) of **Theorem** (5.1.2), $(-1)\boldsymbol{u} = -\boldsymbol{u}$ as required. All the other **axioms (M1) - (M4)** hold because they do in V and this completes the theorem. □

In moment we will give several examples of **vector spaces** and **subspaces**, but before we do we prove one more theorem which concerns **linear combinations**.

Theorem 5.1.4. *Let V be a* **vector space***, W a* **subspace** *of V and $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ a sequence of vectors from W. If $c_1, c_2, \ldots, c_k$ are scalars then the* **linear combination** *$c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \cdots + c_k\boldsymbol{w}_k \in W$.*

Proof. The proof is by **mathematical induction** on k.

Initial steps: The case $k = 1$ is just 2) of **Theorem** (5.1.3). Suppose $k = 2$. We know by 2) of **Theorem** (5.1.3) that $c_1\boldsymbol{w}_1$ and $c_2\boldsymbol{w}_2 \in W$ and then by 1) of the same theorem $c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 \in W$.

Inductive step: Now assume the result is true for any sequence of vectors $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ of length k and scalars $c_1, c_2, \ldots, c_k$ and suppose we have a sequence $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k, \boldsymbol{w}_{k+1})$ in W of length $k+1$ and scalars $c_1, c_2, \ldots, c_k, c_{k+1}$. Set $\boldsymbol{v} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \cdots + c_k\boldsymbol{w}_k \in W$. By the inductive hypothesis, $\boldsymbol{v} \in W$. So now the sequence of $(\boldsymbol{v}, \boldsymbol{w}_{k+1})$ of length two consists of vectors from W. By case of $k = 2$ we know if c and d are any scalars then $c\boldsymbol{v} + d\boldsymbol{w}_{k+1}$ is in W. Taking $c = 1$ and $d = c_{k+1}$ we conclude that $\boldsymbol{v} + c_{k+1}\boldsymbol{w}_{k+1} = [c_1\boldsymbol{w}_1 + \ldots c_k\boldsymbol{w}_k] + c_{k+1}\boldsymbol{w}_{k+1}$ is in W which completes the proof. □

We now proceed to some examples of subspaces.

Example 5.1.6. If V is a **vector space** then V and $\{\mathbf{0}_V\}$ are **subspaces** of V. These are referred to as ***trivial subspaces***. The subspace $\{\mathbf{0}_V\}$ is called the ***zero subspace***.

Example 5.1.7. **Subspaces** of $\mathbb{R}^n$ as previously defined are **subspaces**.

Example 5.1.8. Let n be a **natural number**. Let $\mathbb{R}_n[x]$ be the collection of all polynomials in the indeterminate x which have **degree** at most n. Thus,

$$\mathbb{R}_n[x] = \{a_0 + a_1x + \cdots + a_nx^n \mid a_0, \ldots, a_n \in \mathbb{R}\}.$$

Then $\mathbb{R}_n[x]$ is a **subspace** of $\mathbb{R}[x]$.

Two typical elements of $\mathbb{R}_n[x]$ are: $a_0 + a_1x + \cdots + a_nx^n, b_0 + b_1x + \cdots + b_nx^n$. Their sum is $(a_0 + b_0) + (a_1 + b_1)x + \ldots (a_n + b_n)x^n$ which is in $\mathbb{R}_n[x]$.

Also, for a scalar c, $c(a_0 + a_1x + \cdots + a_nx^n) = (ca_0) + (ca_1)x + \cdots + (ca_n)x^n$ which is also in $\mathbb{R}_n[x]$.

Example 5.1.9. We denote by $C(\mathbb{R})$ the collection of all **continuous functions** from $\mathbb{R}$ to $\mathbb{R}$. This is a **subspace** of $F(\mathbb{R})$. This depends of the following facts proved (stated) in the first calculus class:

The sum of two continuous functions is continuous.

A scalar multiple of a continuous function is continuous.

Example 5.1.10. $D_n(\mathbb{R})$ the collection of all **diagonal matrices** in $M_{n\times n}(\mathbb{R})$ is a **subspace** of $M_{n\times n}(\mathbb{R})$.

Example 5.1.11. $U_n(\mathbb{R})$ the collection of all **upper triangular matrices** in $M_{n\times n}(\mathbb{R})$ is a subspace of $M_{n\times n}(\mathbb{R}$.

Example 5.1.12. Let a be a real number. Set $W = \{f(x) \in \mathbb{R}_n[x] : f(a) = 0\}$. Suppose that $f(x), g(x) \in W$ so that $f(a) = g(a) = 0$. By the definition of $(f+g)(x)$ it follows that $(f+g)(a) = f(a)+g(a) = 0+0 = 0$. So, W is closed under addition. On the other hand, suppose $f \in W$ and c is scalar. We need to show that $cf \in W$ which means we need to show that $(cf)(a) = 0$. However, $(cf)(a) = cf(a) = c \cdot 0 = 0$.

Theorem 5.1.5. *Suppose U and W are **subspaces** of the **vector space** V. Then the **intersection** of U and W, $U \cap W$, is a **subspace** of V.*

Proof. By $U \cap W$ we mean the set consisting of all those vectors which belong to both U and W. Note that $U \cap W$ is nonempty since both U and W contain $\mathbf{0}_V$ and therefore $\mathbf{0}_V \in U \cap W$.

By **Theorem** (5.1.3) we have to show that $U \cap W$ is closed under addition and closed under scalar multiplication.

Suppose $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors in $U \cap W$. Then $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors that are contained in both U and W. Since U is a **subspace** and $\boldsymbol{x}, \boldsymbol{y} \in U$ it follows that $\boldsymbol{x}+\boldsymbol{y} \in U$. Since W is a **subspace** and $\boldsymbol{x}, \boldsymbol{y} \in W$ it follows that $\boldsymbol{x}+\boldsymbol{y} \in W$. Since $\boldsymbol{x}+\boldsymbol{y}$ is in U and in W it is in the intersection and therefore $U+W$ is closed under addition.

Now assume $\boldsymbol{x} \in U \cap W$ and c is a scalar. Since $\boldsymbol{x}$ is in the intersection it is in both U and W. Since it is in U and U is a **subspace** , $c\boldsymbol{x}$ is in U. Since $\boldsymbol{x}$ is in W and W is a **subspace** the scalar multiple $c\boldsymbol{x}$ is in W. Since $c\boldsymbol{x}$ is in U and $c\boldsymbol{x}$ is in W it is in the intersection. Thus, $U \cap W$ is closed under scalar multiplication and therefore a **subspace**. □

Remark 5.2. It is straightforward to prove by induction that if $U_1, \ldots, U_k$ are **subspaces** of a **vector space**, V, then the intersection $U_1 \cap \cdots \cap U_k$ is a subspace of V.

Definition 5.7. Let U and W be **subspaces** of **vector space** V. Define the ***sum*** of U and V, denoted by $U+W$, to be the collection of all vectors which can be written as a sum of a vector $\boldsymbol{u}$ from U and a vector $\boldsymbol{w}$ from W,

$$U+W := \{\boldsymbol{u}+\boldsymbol{w} : \boldsymbol{u} \in U, \boldsymbol{w} \in W\}.$$

In other words, take all sums of something in U and something in W.

Example 5.1.13. Suppose $U = Span(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$ and $W = Span(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_l)$ are **subspaces** of $\mathbb{R}^n$. Then $U+W = Span(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_l)$.

Theorem 5.1.6. *Assume U and W are **subspaces** of a **vector space** V. Then $U+W$ is a **subspace** of V.*

Proof. Suppose $\boldsymbol{x}, \boldsymbol{y} \in U+W$. Then by the definition of the **sum** of **subspaces**, there are elements $\boldsymbol{u}_1 \in U, \boldsymbol{w}_1 \in W$ so $\boldsymbol{x} = \boldsymbol{u}_1 + \boldsymbol{w}_1$ and elements $\boldsymbol{u}_2 \in U, \boldsymbol{w}_2 \in W$ so that $\boldsymbol{y} = \boldsymbol{u}_2 + \boldsymbol{w}_2$. Then

$$\boldsymbol{x}+\boldsymbol{y} = (\boldsymbol{u}_1+\boldsymbol{w}_1)+(\boldsymbol{u}_2+\boldsymbol{w}_2) = (\boldsymbol{u}_1+\boldsymbol{u}_2)+(\boldsymbol{w}_1+\boldsymbol{w}_2).$$

Since U is a **subspace** of V, $\boldsymbol{u}_1+\boldsymbol{u}_2 \in U$. Likewise, since W is a **subspace** of V $\boldsymbol{w}_1+\boldsymbol{w}_2 \in W$. Therefore $\boldsymbol{x}+\boldsymbol{y} = (\boldsymbol{u}_1+\boldsymbol{u}_2)+(\boldsymbol{w}_1+\boldsymbol{w}_2) \in U+W$. So $U+W$ is **closed under addition**.

We leave the second part, that $U + W$ is **closed under scalar multiplication** as a **challenge exercise**. $\square$

What You Can Now Do

1. Determine whether a given subset of a **vector space** is a **subspace**.

Methods (How To Do It)

Method 5.1.1. Determine whether a given subset of a **vector space** is a **subspace**.

The typical exercise will describe the subset in terms of some kind of formulas. As a rule of thumb, these formulas should be linear expressions without constant term. We do a few examples to illustrate.

Example 5.1.14. Let $m(x,y) = \begin{pmatrix} 2x+3y & x-2y \\ -x+3y & 4x+y \end{pmatrix}$ and set $W = \{m(x,y) : x, y \in \mathbb{R}\}$. Determine if W is a **subspace** of $M_{2\times 2}(\mathbb{R})$ and justify your conclusion.

Since the expression for the entries of the matrix are linear expressions without constant term it should be a **subspace**. We need to show that it is **closed under addition and scalar multiplication**.

Typical elements are

$$m(x_1,y_1) = \begin{pmatrix} 2x_1+3y_1 & x_1-2y_1 \\ -x_1+3y_1 & 4x_1+y_1 \end{pmatrix}, m(x_2,y_2) = \begin{pmatrix} 2x_2+3y_2 & x_2-2y_2 \\ -x_2+3y_2 & 4x_2+y_2 \end{pmatrix}.$$

We need to show that $m(x_1,y_1) + m(x_2,y_2) \in W$.

However,

$$m(x_1,y_1)+m(x_2,y_2) = \begin{pmatrix} (2x_1+3y_1)+(2x_2+3y_2) & (x_1-2y_1)+(x_2-2y_2) \\ (-x_1+3y_1)+(-x_2+3y_2) & (4x_1+y_1)+(4x_2+y_2) \end{pmatrix} =$$

$$\begin{pmatrix} 2(x_1+x_2)+3(y_1+y_2) & (x_1+x_2)-2(y_1+y_2) \\ -(x_1+x_2)+3(y_1+y_2) & 4(x_1+x_2)+(y_1+y_2) \end{pmatrix} = m(x_1+x_2, y_1+y_2)$$

which is an element of W.

We also have to show that for a scalar c that $cm(x,y) \in W$.

$$cm(x,y) = c\begin{pmatrix} 2x+3y & x-2y \\ -x+3y & 4x+y \end{pmatrix} = \begin{pmatrix} c(2x+3y) & c(x-2y) \\ c(-x+3y) & c(4x+y) \end{pmatrix} =$$

$\begin{pmatrix} 2(cx)+3(cy) & cx-2(cy) \\ -cx+3(cy) & 4(cx)+cy \end{pmatrix} = m(cx,cy) \in W$. So W is a **subspace** .

If the components are given by linear expressions but there are constant terms it could be a **subspace** but often won't be. Try and show that the **zero vector** is not included. This will reduce to solving a **linear system** . If the zero vector is not an element then it is not a **subspace**. If it is then it will be a subspace.

Example 5.1.15. Let $m(x,y) = \begin{pmatrix} 3x+4y+1 & 2x+3y+1 \\ y+1 & x+y \end{pmatrix}$ and set

$$W = \{m(x,y) : x, y \in \mathbb{R}\}.$$

Determine whether W is a **subspace** of $M_{2\times 2}(\mathbb{R})$ and justify your conclusion.

We first examine whether the **zero vector** of $M_{2\times 2}(\mathbb{R})$, $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. is in W.

So we have to determine whether there are real numbers x, y such that

$$\begin{array}{ccccccc} 3x & + & 4y & + & 1 & = & 0 \\ 2x & + & 3y & + & 1 & = & 0 \\ & & y & + & 1 & = & 0 \\ x & + & y & & & = & 0 \end{array}$$

This **linear system** is **equivalent** to the following **linear system** in which the equations are in standard form:

$$\begin{array}{ccccc} 3x & + & 4y & = & -1 \\ 2x & + & 3y & = & -1 \\ & & y & = & -1 \\ x & + & y & = & 0 \end{array}$$

To solve this **linear system** we write the **augmented matrix** and use **Gaussian elimination** to obtain an **reduced echelon form**:

$$\begin{pmatrix} 3 & 4 & | & -1 \\ 2 & 3 & | & -1 \\ 0 & 1 & | & -1 \\ 1 & 1 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & | & 0 \\ 2 & 3 & | & -1 \\ 0 & 1 & | & -1 \\ 3 & 4 & | & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & | & 0 \\ 0 & 1 & | & -1 \\ 0 & 1 & | & -1 \\ 0 & 1 & | & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & | & 1 \\ 0 & 1 & | & -1 \\ 0 & 0 & | & 0 \\ 0 & 0 & | & 0 \end{pmatrix}.$$

Since the last column (the augmeted column) is not a **pivot column**, the **linear system** is **consistent** and the **zero vector** belongs to W. This will mean that W is a **subspace** . Actually demonstrating this at this point takes a bit of computation. For example one has to show that given $(a,b),(c,d)$ there is an (x,y) such that $m(a,b)+m(c,d) = m(x,y)$, that is,

$$\begin{pmatrix} 3a+4b+1 & 2a+3b+1 \\ b+1 & a+b \end{pmatrix} + \begin{pmatrix} 3c+4d+1 & 2c+3d+1 \\ d+1 & c+d \end{pmatrix} =$$

$$\begin{pmatrix} 3x+4y+1 & 2x+3y+1 \\ y+1 & x+y \end{pmatrix}.$$

This gives rise to the following **linear system**

$$\begin{array}{ccccccc} 3x & + & 4y & + & 1 & = & 3a+4b+3c+4d+2 \\ 2x & + & 3y & + & 1 & = & 2a+3b+2c+3d+2 \\ & & y & + & 1 & = & b+d+2 \\ x & + & y & & & = & a+b+c+d \end{array} \tag{5.13}$$

which is **equivalent to** the linear system

$$\begin{array}{ccccc} 3x & + & 4y & = & 3a+4b+3c+4d+1 \\ 2x & + & 3y & = & 2a+3b+2c+3d+1 \\ & & y & = & b+d+1 \\ x & + & y & = & a+b+c+d \end{array}.$$

We solve by forming the **augmented matrix** and use **Gaussian elimination** to obtain an **echelon form**.

$$\left(\begin{array}{cc|c} 3 & 4 & 3a+4b+3c+4d+1 \\ 2 & 3 & 2a+3b+2c+3d+1 \\ 0 & 1 & b+d+1 \\ 1 & 1 & a+b+c+d \end{array}\right) \to$$

$$\left(\begin{array}{cc|c} 1 & 1 & a+b+c+d \\ 2 & 3 & 2a+3b+2c+3d+1 \\ 0 & 1 & b+d+1 \\ 3 & 4 & 3a+4b+3c+4d+1 \end{array}\right) \to$$

$$\left(\begin{array}{cc|c} 1 & 1 & a+b+c+d \\ 0 & 1 & b+d+1 \\ 0 & 1 & b+d+1 \\ 0 & 1 & b+d+1 \end{array}\right) \to$$

$$\left(\begin{array}{cc|c} 1 & 0 & a+c-1 \\ 0 & 1 & b+d+1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

Since the last column (augmented column) is not a **pivot column** we conclude that the **linear system** (5.13) is **consistent** and W is **closed under addition**. Scalar multiplication is done similarly.

Another way to define a **subspace** is to put some conditions on the "components" of the vectors. For example, we could consider all elements $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$ such that $f(a_{11}, a_{12}, a_{21}, a_{22}) = 0$. Here the rule is hard and fast, it will be a **subspace** if and only if the expression on the left is linear without a constant term.

Example 5.1.16. Let $W = \{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in M_{22} : a_{11} + a_{12} + a_{21} - 3a_{22} = 0\}$.

This is a **subspace** . For suppose $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \in W$. We need to show that

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix} \in W.$$

This means we need to show that

$(a_{11} + b_{11}) + (a_{12} + b_{12}) + (a_{21} + b_{21}) - 3(a_{22} + b_{22}) = 0.$

However, $(a_{11} + b_{11}) + (a_{12} + b_{12}) + (a_{21} + b_{21}) - 3(a_{22} + b_{22}) =$

$[a_{11} + a_{12} + a_{21} - 3a_{22}] + [b_{11} + b_{12} + b_{21} - 3b_{22}] = 0 + 0 = 0$ as required.

Scalar multiplication is done similarly.

On the other hand, if the expression is linear but the constant is not zero, for example $2a_{11} - 3a_{12} + 4a_{21} - 5a_{22} - 6 = 0$ then the zero vector does not satisfy the condition and so it is not a **subspace** . Also, if the expression is not linear then it will generally not be closed under scalar multiplication and therefore is not a **subspace**.

Exercises

For all these exercises see **Method** (5.1.1).

In 1 - 4 determine whether the subset $W = \{m(x, y) : x, y \in \mathbb{R}\}$is a **subspace** of $M_{2\times 2}(\mathbb{R})$ for the given $m(x, y)$.

1. $m(x, y) = \begin{pmatrix} 2x + y & 5x + 2y \\ -x + y & x + y \end{pmatrix}$.

2. $m(x, y) = \begin{pmatrix} x + 3y & 2x + 7y \\ 3x + 2y & 4x + 3y \end{pmatrix}$.

3. $m(x, y) = \begin{pmatrix} x + y + 1 & x + 2y \\ 3x - y & 2x + 5y \end{pmatrix}$.

4. $m(x, y) = \begin{pmatrix} x^2 - y^2 & x + y \\ x - y & xy \end{pmatrix}$.

In 5 - 7 the subset W is given by $W = \{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} : f(a_{11}, a_{12}, a_{21}, a_{22}) = 0\}$. Determine if W is a **subspace** of $M_{2\times 2}(\mathbb{R})$ if the expression f is as given.

5. $f(a_{11}, a_{12}, a_{21}, a_{22}) = a_{11} - a_{12} + a_{21} - a_{22}$.

6. $f(a_{11}, a_{12}, a_{21}, a_{22}) = a_{11} + a_{22}$.

7. $f(a_{11}, a_{12}, a_{21}, a_{22}) = a_{12} - a_{21} - 1$.

8. $f(a_{11}, a_{12}, a_{21}, a_{22}) = a_{11}a_{22}$.

In 9 - 12 determine whether the subset $W = \{f(a, b) : a, b \in \mathbb{R}\}$ is a **subspace** of $\mathbb{R}_2[x]$ for the given $f(a, b)$.

9. $f(a, b) = (a - b) + (2a + b)X + (3a + 4b)X^2$.

10. $f(a, b) = (2a - 3b + 1) + (-2a + 5b)X + (2a + b)X^2$.

11. $f(a, b) = ab + (a - b)X + (a + b)X^2$.

12. $f(a, b) = (b + 1) + (a + 2b + 1)X + (2a + 3b + 1)X^2$.

In 13 - 17 the subset W is given by $W = \{a_0 + a_1X + \cdots + a_nX^n : f(a_0, a_1, \ldots, a_n) = 0\}$. Determine if W is a **subspace** of $\mathbb{R}_n[x]$ if the expression f is as given.

13. $f(a_0, a_1, \ldots, a_n) = a_0 + a_1 + \cdots + a_n$.

14. $f(a_0, a_1, \ldots, a_n) = a_0$.

15. $f(a_0, a_1, \ldots, a_n) = a_1$.

16. $f(a_0, a_1, \ldots, a_n) = a_0 - 1$.

17. $f(a_0, a_1, \ldots, a_n) = a_0a_1$.

In exercises 18 - 22 answer true or false and give an explanation.

18. $\mathbb{R}^3$ is a **subspace** of $\mathbb{R}^4$.

19. $\mathbb{R}_3[x]$ is a **subspace** of $\mathbb{R}_4[x]$.

20. If V is a **vector space** and V contains an element $\boldsymbol{v} \neq \mathbf{0}_V$ then V has infinitely many elements.

21. If V is a **vector space** and W is a **subspace** then the **zero vector** of V, $\mathbf{0}_V$, belongs to W.

22. If V is a **vector space**, W a **subspace** of V, and U a subspace of W then U is a subspace of V.

Challenge Exercises (Problems)

1. Let V be a **vector space** with **additive identity** $\mathbf{0}_V$ and let c be a scalar. Prove that $c\mathbf{0}_V = \mathbf{0}_V$. Hint: Use the fact that $\mathbf{0}_V + \mathbf{0}_V = \mathbf{0}_V$ and use the **distributive property**.

2. Let V be a **vector space**, U, W **subspaces** and $U+W := \{u+w : u \in U, w \in W\}$. Prove that $U + W$ is **closed under scalar multiplication.**

3. Let V be a **vector space**. Prove the following cancellation property: Assume $\boldsymbol{v}, \boldsymbol{x}$, and $\boldsymbol{y}$ are vectors in V and $\boldsymbol{v} + \boldsymbol{x} = \boldsymbol{v} + \boldsymbol{y}$. Then $\boldsymbol{x} = \boldsymbol{y}$.

4. Let V be a **vector space**. Prove the following cancelation property: Assume $c \neq 0$ is scalar and $c\boldsymbol{x} = c\boldsymbol{y}$ then $\boldsymbol{x} = \boldsymbol{y}$.

Quiz Solutions

1. There is no choice of (a, b, c) such that $m(a, b, c) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. In order for **two matrices to be equal** they must have equal entries. This leads to the **linear system**

$$\begin{array}{ccccccc} 2a & - & 3b & + & c & = & 1 \\ 2a & - & 4b & + & 2c & = & 0 \\ a & + & b & + & c & = & 0 \\ 3a & - & 2b & + & 2c & = & 0 \end{array}$$

The **augmented matrix** of this **linear system** is $\left(\begin{array}{ccc|c} 2 & -3 & 1 & 1 \\ 2 & -4 & 2 & 0 \\ 1 & 1 & 1 & 0 \\ 3 & -2 & 2 & 0 \end{array}\right)$ and has an **echelon form** $\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$. So the last column is a **pivot column** and there are no **solutions**.

Not right, see **Method** (1.2.4)

2. There is a unique choice of (a, b, c) such that $m(a, b, c) = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$, namely, $(a, b, c) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. As in 1 we get a **linear system** , this time with **augmented matrix** $\left(\begin{array}{ccc|c} 2 & -3 & 1 & 0 \\ 2 & -4 & 2 & 0 \\ 1 & 1 & 1 & 1 \\ 3 & -2 & 2 & 1 \end{array}\right)$ which has **reduced echelon form** $\left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{array}\right)$.

Not right, see **Method** (1.2.4)

3. There is a unique **solution**, $(a, b, c) = (-\frac{4}{3}, -\frac{1}{3}, 1)$. This exercise leads to the **linear system**

$$\begin{array}{ccccccc} a & + & 2b & + & c & = & -1 \\ 2a & + & b & + & 3c & = & 0 \\ a & + & 2b & + & 2c & = & 0 \\ 2a & + & b & + & 3c & = & 0 \end{array}$$

The **augmented matrix** $\begin{pmatrix} 1 & 2 & 1 & | & -1 \\ 2 & 1 & 3 & | & 0 \\ 1 & 2 & 2 & | & 0 \\ 2 & 1 & 3 & | & 0 \end{pmatrix}$ has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 0 & | & -\frac{4}{3} \\ 0 & 1 & 0 & | & -\frac{1}{3} \\ 0 & 0 & 1 & | & 1 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$.

Not right, see **Method** (1.2.4)

5.2. Span and Independence in Vector Spaces

The concepts of span, linear independence and basis are introduced for abstract **vector spaces**.

Am I Ready for This Material

Extensive use is made of the following concepts and procedures in this section and they need to be mastered to fully understand the new ideas and methods introduced here:

matrix in echelon form

matrix in reduced echelon form

an echelon form of a matrix

the reduced echelon form of a matrix

pivot position of a matrix

pivot column of a matrix

span of a sequence of vectors in $\mathbb{R}^n$

spanning sequence

linearly dependent sequence of vectors in $\mathbb{R}^n$

linearly independent sequence of vectors in $\mathbb{R}^n$

basis of a subspace of $\mathbb{R}^n$

null space of a matrix

vector space

subspace of a vector space

You will need to know the following methods:

Gaussian elimination

procedure for solving a linear system

Quiz

1. Verify that the sequence of vectors $\left(\begin{pmatrix}-1\\2\\5\end{pmatrix}, \begin{pmatrix}5\\2\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right)$ is **linearly dependent** and find a **non-trivial dependence relation**.

Let $\boldsymbol{v}_1 = \begin{pmatrix}1\\0\\-1\end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix}1\\1\\2\end{pmatrix}$ and $\boldsymbol{v}_3 = \begin{pmatrix}2\\1\\2\end{pmatrix}$.

2. Verify that the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**.

3. Why can we conclude that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = \mathbb{R}^3$?

4. Making use of 2) and the **definition of linear independence** give an argument for why the sequence of vectors $(\boldsymbol{v}_1', \boldsymbol{v}_2', \boldsymbol{v}_3')$ is **linearly independent** where

$$\boldsymbol{v}_1' = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \boldsymbol{v}_2' = \begin{pmatrix} 1 \\ 1 \\ 2 \\ -4 \end{pmatrix}, \boldsymbol{v}_3' = \begin{pmatrix} 2 \\ 1 \\ 2 \\ -5 \end{pmatrix}.$$

Let $S = \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} : x_1 + x_2 + x_3 + x_4 = 0\}$.

5. Show that S is a **subspace** of $\mathbb{R}^4$.

6. Prove that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a **basis** of S.

Quiz Solutions

New Concepts

In this section we generalize the most basic concepts that we have introduced in the context of $\mathbb{R}^n$ to **abstract vectors spaces**. These concepts are

span of a sequence of vectors from a vector space

spanning sequence of a vector space V

spanning sequence of a subspace of a vector space

a sequence of vectors in a vector space is linearly dependent

a sequence of vectors in a vector space is linearly independent

basis of a subspace of a vector space

Theory (Why It Works)

We generalize the concepts of span, independence and basis, which were defined for $\mathbb{R}^n$, to abstract **vector spaces**.

Definition 5.8. Let V be a vector space and $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be a sequence of vectors from V. The set of all linear combinations of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is called the ***span*** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ and is denoted by $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.

More generally, if S is any subset of V then by $Span(S)$ we mean the collection of all vectors $\boldsymbol{u}$ for which there exists a sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of vectors from S and scalars $c_1, \ldots, c_k$ such that $\boldsymbol{u} = c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k$. In other words, $Span(S)$ is the union of all $Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ such that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is a sequence of vectors from S.

We will be interested in sequences and subsets S of a vector space V such that $V = Span(S)$. These are the subject of our next definition:

Definition 5.9. Let V be a vector space and W a subspace of V. If $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k)$ is a sequence of vectors from W and $Span(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k) = W$ then $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k)$ is a ***spanning sequence*** of W. We also say that $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k)$ ***spans*** W.

Sometimes it is not possible to find a sequence of vectors that span a space, for example, there is no finite sequence of vectors which spans the space $\mathbb{R}[x]$ (this is left as a challenge exercise). For this reason we introduce the following notion:

If S is a subset of the subspace W and $Span(S) = W$ then we say that S is a ***spanning set*** of W and that S ***spans*** W. Note that, unlike a sequence, a set cannot have repeated vectors.

Convention

The span of the empty sequence is the zero subspace $\{\boldsymbol{0}_V\}$.

The first important question that arises is can we determine if some vector $\boldsymbol{u}$ in a vector space V is in the span of a given sequence of vectors, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$. The following example illustrates how to answer such a question.

Example 5.2.1. Determine if the polynomial $2 + 3x - 5x^2$ is in the span of $(\boldsymbol{f}, \boldsymbol{g})$ where $\boldsymbol{f}(x) = 1 + x - 2x^2$ and $\boldsymbol{g}(x) = 2 + x - 3x^2$.

We need to determine if there are scalars a, b such that

$$a\boldsymbol{f}(x) + b\boldsymbol{g}(x) = a(1 + x - 2x^2) + b(2 + x - 3x^2) = 2 + 3x - 5x^2.$$

As we shall see, this reduces to one of our standard problems, namely whether or not a linear system is consistent. We express the right hand side as a polynomial with coefficients in terms of a and b :

$$(a+2b)+(a+b)x+(-2a-3b)x^2 = 2+3x-5x^2.$$

If $(a+2b)+(a+b)x+(-2a-3b)x^2 = 2+3x-5x^2 = 2+3x-5x^2$ then the constants terms on both sides must be equal, $a+2b=2$; the coefficient of x in both expressions must be equal and so $a+b=3$; and the coefficients of x^2 must be equal, yielding $-2a-3b=-5$. This gives rise to the **linear system**:

$$\begin{array}{ccccc} a & + & 2b & = & 2 \\ a & + & b & = & 3 \\ -2a & + & -3b & = & -5 \end{array}.$$

This **linear system** has **augmenented matrix**

$$\begin{pmatrix} 1 & 2 & | & 2 \\ 1 & 1 & | & 3 \\ -2 & -3 & | & -5 \end{pmatrix}$$

Thus, $2+3x-5x^2 \in Span(\boldsymbol{f}(x), \boldsymbol{g}(x))$ if and only if the **3-vector** $\begin{pmatrix} 2 \\ 3 \\ -5 \end{pmatrix}$ is in $Span\left(\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}\right)$.

Note that the first column has as its entries the coefficients of $1, x, x^2$ in $\boldsymbol{f}$, the second column the coefficients of $1, x, x^2$ in $\boldsymbol{g}$ and the third column the coefficients of $1, x, x^2$ in $2+3x-5x^2$. So, we can think of this matrix as arising in the following way: we associate to an element $a_0 + a_1x + a_2x^2$ the column vector $\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$. Then the **augmented matrix** is made from these columns.

We use **Gaussian elimination** to obtain the **reduced echelon form** of the **augmented matrix** and can then determine if the **linear system** is **consistent**.

$$\to \begin{pmatrix} 1 & 2 & | & 2 \\ 0 & -1 & | & 1 \\ 0 & 1 & | & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & | & 4 \\ 0 & 1 & | & -1 \\ 0 & 0 & | & 0 \end{pmatrix}$$

Thus, there is a unique solution, $a = 4, b = -1$ and therefore

$$2+3x-5x^2 \in Span(1+x-2x^2, 2+x-3x^2).$$

Example 5.2.2. More generally, suppose $\boldsymbol{f}_1(x) = a_{01} + a_{11}x + a_{21}x^2 + \cdots + a_{n1}x^n, \boldsymbol{f}_2(x) = a_{02} + a_{12}x + a_{22}x^2 + \cdots + a_{n2}x^n, \ldots, \boldsymbol{f}_k(x) = a_{0k} + a_{1k}x + a_{2k}x^2 + \cdots + a_{nk}x^n$ and $h(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n$ are polynomials in $\mathbb{R}_n[x]$ and we want to determine if $\boldsymbol{g}(x) \in Span(\boldsymbol{f}_1(x), \boldsymbol{f}_2(x), \ldots, \boldsymbol{f}_k(x))$.

We have to determine if there are scalars $c_1, c_2, \ldots, c_k$ such that

$$\boldsymbol{g}(x) = c_1\boldsymbol{f}_1(x) + c_2\boldsymbol{f}_2(x) + \cdots + c_k\boldsymbol{f}_k(x)$$

$$b_0 + b_1x + b_2x^2 + \cdots + b_nx^n = c_1(a_{01} + a_{11}x + a_{21}x^2 + \cdots + a_{n1}x^n) +$$

$$c_2(a_{02} + a_{12}x + a_{22}x^2 + \cdots + a_{n2}x^n) + \cdots +$$

$$c_k(a_{0k} + a_{1k}x + a_{2k}x^2 + \cdots + a_{nk}x^n) \tag{5.14}$$

We can distribute on the right hand side of (5.14) and then collection terms to get the constant term, the coefficient of x, of x^2 and so on. When we do this we obtain the following expression:

$$(c_1a_{01} + c_2a_{02} + \ldots c_ka_{0k}) + (c_1a_{11} + c_2a_{12} + \ldots c_ka_{1k})x+$$

$$(c_1a_{21} + c_2a_{22} + \ldots c_ka_{2k})x^2 + \cdots + (c_1a_{n1} + c_2a_{n2} + \ldots c_ka_{nk})x^n \tag{5.15}$$

Since two polynomials are equal if and only if all their coefficients are equal we can conclude that $\boldsymbol{g}(x)$ is equal to $c_1\boldsymbol{f}_1(x) + c_2\boldsymbol{f}_2(x) + \cdots + c_k\boldsymbol{f}_k(x)$ if and only if $\begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$ is a **solution** to the **linear system**

$$\begin{aligned} a_{01}x_1 + a_{02}x_2 + \ldots a_{0k}x_k &= b_1 \\ a_{11}x_1 + a_{12}x_2 + \ldots a_{1k}x_k &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots a_{nk}x_k &= b_n \end{aligned}$$

This is an **inhomogeneous linear system** and it has **augmented matrix**

$$\begin{pmatrix} a_{01} & a_{02} & \dots & a_{0k} & | & b_0 \\ a_{11} & a_{12} & \dots & a_{1k} & | & b_1 \\ \vdots & \vdots & \dots & \vdots & | & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} & | & b_n \end{pmatrix} \tag{5.16}$$

Note that the first column of this matrix consists of the coefficient of $1, x, \dots, x^n$ in $\boldsymbol{f}_1$, the second column consists of the coefficients of $1, x, \dots, x^n$ in $\boldsymbol{f}_2, \dots,$ and, in general, for $1 \leq j \leq k$ the j^{th} column consists of the coefficients of $1, x, \dots, x^n$ in $\boldsymbol{f}_j$. Finally, the last (augmented) column consists of the coefficients of $1, x, \dots, x^n$ in $\boldsymbol{g}(x)$.

Example 5.2.3. Let $\boldsymbol{m}_1 = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 1 & 3 \\ -2 & -2 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 1 & -4 \\ 2 & 1 \end{pmatrix}$ and $\boldsymbol{m} = \begin{pmatrix} 3 & -6 \\ 2 & 1 \end{pmatrix}$. Determine if $\boldsymbol{m} \in Span(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$.

We have to determine if there are scalars c_1, c_2, c_3 such that

$$c_1\boldsymbol{m}_1 + c_2\boldsymbol{m}_2 + c_3\boldsymbol{m}_3 = \boldsymbol{m}.$$

Substituting the actual matrices this becomes

$$c_1 \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} + c_2 \begin{pmatrix} 1 & 3 \\ -2 & -2 \end{pmatrix} + c_3 \begin{pmatrix} 1 & -4 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -6 \\ 2 & 1 \end{pmatrix}. \tag{5.17}$$

After performing the scalar multiplications and matrix additions in (5.17) we obtain

$$\begin{pmatrix} c_1 + c_2 + c_3 & c_1 + 3c_2 - 4c_3 \\ -c_1 - 2c_2 + 2c_3 & -c_1 - 2c_2 + c_2 \end{pmatrix} = \begin{pmatrix} 3 & -6 \\ 2 & 1 \end{pmatrix}. \tag{5.18}$$

Note that these two matrices are equal if and only if their components are all equal. Setting these equal gives rise to an **inhomogeneous linear system** of four equations (one for each component) in three variables (c_1, c_2, c_3). This system is:

$$\begin{array}{rcrcrcr} c_1 & + & c_2 & + & c_3 & = & 3 \\ -c_1 & - & 2c_2 & + & 2c_3 & = & 2 \\ c_1 & + & 3c_2 & - & 4c_3 & = & -6 \\ -c_1 & - & 2c_2 & + & c_2 & = & 1 \end{array} \tag{5.19}$$

This **linear system** has **augmented matrix**

$$\begin{pmatrix} 1 & 1 & 1 & | & 3 \\ -1 & -2 & 2 & | & 2 \\ 1 & 3 & -4 & | & -6 \\ -1 & -2 & 1 & | & 1 \end{pmatrix}. \tag{5.20}$$

Note the relationship of the columns of this matrix to the matrices $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3$ and $\boldsymbol{m}$. The first column consists of the components of $\boldsymbol{m}_1$ the second of $\boldsymbol{m}_2$ the third with those of $\boldsymbol{m}_3$ and the augmented column contains the components of $\boldsymbol{m}$. These columns are obtained, in general, by taking a matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ to the **4-vector** $\begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{pmatrix}$.

The **reduced echelon form** of the augmented matrix in (5.20), obtained by **Gaussian elimination**, is

$$\begin{pmatrix} 1 & 0 & 0 & | & 4 \\ 0 & 1 & 0 & | & -2 \\ 0 & 0 & 1 & | & 1 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \tag{5.21}$$

We conclude from this that $\boldsymbol{m}$ is a **linear combination** of the sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$ and, in fact, $\boldsymbol{m} = 4\boldsymbol{m}_1 - 2\boldsymbol{m}_2 + \boldsymbol{m}_3$.

Example 5.2.4. Recall that the matrix $E_{ij}^{m\times n}$ is the $m \times n$ matrix with a 1 in the $(i, j)-$entry and 0's everywhere else. So, for example

$$E_{12}^{2\times 3} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In general $M_{m\times n}(\mathbb{R})$ is spanned by the sequence $(E_{11}, E_{21}, \ldots, E_{m1}, E_{12}, E_{22}, \ldots, E_{m2}, \ldots, E_{1n}, E_{2n}, \ldots, E_{mn})$ (here we have dropped the superscript $m \times n$ because it is clear from the context).

More concretely, if $E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ then $(E_{11}, E_{21}, E_{12}, E_{22})$ **spans** $M_{2\times 2}(\mathbb{R})$.

Example 5.2.5. Let $S = \{1\} \cup \{x^k | k = 1, 2, \ldots\}$. Then S is a **spanning set** of $\mathbb{R}[x]$.

Example 5.2.6. $\mathbb{R}_n[x] = Span(1, x, \ldots, x^n)$.

Example 5.2.7. Let $\boldsymbol{f}_1(x) = \frac{(x-2)(x-3)}{2}$, $\boldsymbol{f}_2(x) = \frac{(x-1)(x-3)}{-1}$, $\boldsymbol{f}_3(x) = \frac{(x-1)(x-2)}{2}$. Let's compute $f_{ij} = \boldsymbol{f}_i(j)$ for $i = 1, 2, 3; j = 1, 2, 3$.

$$f_{11} = \boldsymbol{f}_1(1) = \frac{(-1)(-2)}{2} = 1 \quad f_{12} = \boldsymbol{f}_1(2) = \frac{(0)(-1)}{2} = 0$$

$$f_{13} = \boldsymbol{f}_1(3) = \frac{(1)(0)}{2} = 0 \quad f_{21} = \boldsymbol{f}_2(1) = \frac{(0)(-2)}{-1} = 0$$

$$f_{22} = \boldsymbol{f}_2(2) = \frac{(1)(-1)}{-1} = 1 \quad f_{23} = \boldsymbol{f}_2(3) = \frac{(2)(0)}{-1} = 0$$

$$f_{31} = \boldsymbol{f}_3(1) = \frac{(0)(-1)}{2} = 0 \quad f_{32} = \boldsymbol{f}_3(2) = \frac{(1)(0)}{2} = 0$$

$$f_{33} = \boldsymbol{f}_3(3) = \frac{(2)(1)}{2} = 1.$$

Therefore the matrix $(f_{ij}) = I_3$, that is, the 3×3 **identity matrix**.

We claim that $Span(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3) = \mathbb{R}_2[x]$. In fact, if $\boldsymbol{f}(x)$ is a polynomial in $\mathbb{R}_2[x]$ and $\boldsymbol{f}(1) = a, \boldsymbol{f}(2) = b, \boldsymbol{f}(3) = c$ then $\boldsymbol{f}(x) = a\boldsymbol{f}_1(x) + b\boldsymbol{f}_2(x) + c\boldsymbol{f}_3(x)$.

This follows since $\boldsymbol{f}(x) - (a\boldsymbol{f}_1(x) + b\boldsymbol{f}_2(x) + c\boldsymbol{f}_3(x))$ has degree at most two but vanishes (is equal to zero) at $x = 1, 2, 3$. However, a nonzero polynomial of degree at most two can have no more than two roots. Therefore $\boldsymbol{f}(x) - (a\boldsymbol{f}_1(x) + b\boldsymbol{f}_2(x) + c\boldsymbol{f}_3(x))$ must be the zero polynomial which means that $\boldsymbol{f}(x) = a\boldsymbol{f}_1(x) + b\boldsymbol{f}_2(x) + c\boldsymbol{f}_3(x)$, a **linear combination** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$.

We now prove a general result:

Theorem 5.2.1. *Let S be either a sequence or a set from a* **vector space** *V. Then the following hold:*

1. $Span(S)$ is a **subspace** *of V.*

2. If W is a **subspace** *of V and W contains S then W contains $Span(S)$.*

Before preceding to the proof we make some necessary definitions and illustrate some of the techniques we use in the proof.

Definition 5.10. Let $A = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k)$ and $B = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_l)$ be two sequences of vectors in a **vector space** V. By the ***union of the two sequences*** A and B we mean the sequence obtained by putting the vectors of B after the vectors in A and denote this by $A \cup B$. Thus, $A \cup B = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_l)$.

Lemma 5.2.2. *Let $A = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$ and $B = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_l)$ be sequences from the* **vector space** *V. Then any vector in $Span(A)$ or $Span(B)$ is in $Span(A \cup B)$.*

Proof. To see this, suppose $\boldsymbol{x} = a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \cdots + a_k\boldsymbol{u}_k$. Then $\boldsymbol{x} = a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \cdots + a_k\boldsymbol{u}_k + 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_l$ which is a **linear combination** of $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_l)$.

Similarly, if $\boldsymbol{y} = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_m\boldsymbol{v}_l$ then $\boldsymbol{y} = 0\boldsymbol{u}_1 + 0\boldsymbol{u}_2 + \cdots + 0\boldsymbol{u}_k + b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_m\boldsymbol{v}_l \in Span(A \cup B)$.

Thus, $Span(A), Span(B) \subset Span(A \cup B)$. □

Proof of **Theorem** (5.2.1)

1. If S is a sequence then the proof is virtually the same as **Theorem** (2.3.5) and so we omit it. We therefore proceed to the case that S is an infinite subset of V. We need to show that $Span(S)$ is i) **closed under addition**; and ii) **closed under scalar multiplication**.

i) Assume $\boldsymbol{x}, \boldsymbol{y} \in Span(S)$. Then there are sequences $A = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ and $B = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_l)$ of vectors from S such that $\boldsymbol{x} \in Span(A)$ and $\boldsymbol{y} \in Span(B)$. By **Lemma** (5.2.2) $\boldsymbol{x}$ and $\boldsymbol{y}$ are contained in $Span(A \cup B)$. Since $A \cup B$ is a sequence of vectors from S it follows that $\boldsymbol{x} + \boldsymbol{y} \in Span(S)$.

ii) Assume $\boldsymbol{x} \in Span(S)$ and c is a scalar. Then there is a sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of vectors from S such that $\boldsymbol{x} \in Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$. By the finite case of the theorem, $c\boldsymbol{x} \in Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) \subset Span(S)$.

2. This follows immediately from **Theorem** (5.1.4). □

Remark 5.3. The two parts of **Theorem** (5.2.1) imply that $Span(S)$ is the "smallest" subspace of V which contains S, that is, if W is a subspace containing S and $W \subset Span(S)$ then $W = Span(S)$.

A corollary of **Theorem** (5.2.1) is:

Corollary 5.2.3. *1. If W is a subspace of a* **vector space** *V then*

$$Span(W) = W.$$

2. If S is a subset of a **vector space** *V then $Span(Span(S)) = Span(S)$.*

Most of the results of Section (2.3) carry over to arbitrary **vector spaces**. We state these for future reference but omit proofs since they are virtually the same as in Section (2.3).

Theorem 5.2.4. *If one of the vectors of the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is a* **linear combination** *of the remaining vectors then that vector can be deleted from the sequence and the* **span** *remains the same.*

We now turn our attention to the concept of linear dependence and linear independence in an abstract **vector space**.

Definition 5.11. A finite sequence of vectors, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ from a **vector space** V is ***linearly dependent*** if there are scalars $c_1, c_2, \ldots, c_k$, not all zero, such that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = 0$. Such an expression is called a ***non-trivial dependence relation***.

More generally, a subset S of a **vector space** V is ***linearly dependent*** if there exists a sequence of **distinct** vectors $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ from S such that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is linearly dependent.

Definition 5.12. A sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ from a **vector space** V said to be ***linearly independent*** if it is not linearly dependent. This means if $c_1, c_2, \ldots, c_k$ are scalars such that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = 0$ then $c_1 = c_2 = \cdots = c_k = 0$. Put succinctly, the only dependence relation is the ***trivial one***.

Finally, a subset S of a **vector space** V is ***linearly independent*** if every sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of distinct vectors from S is linearly independent.

Example 5.2.8. The sequence

$$\left(\begin{pmatrix}2 & 4\\ -5 & -1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \begin{pmatrix}0 & 1\\ 0 & -1\end{pmatrix}, \begin{pmatrix}0 & 0\\ 1 & -1\end{pmatrix}\right)$$

is **linearly dependent** in $M_{2\times 2}(\mathbb{R})$ since

$$\begin{pmatrix}2 & 4\\ -5 & -1\end{pmatrix} + (-2)\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} + (-4)\begin{pmatrix}0 & 1\\ 0 & -1\end{pmatrix} + 5\begin{pmatrix}0 & 0\\ 1 & -1\end{pmatrix} = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}.$$

Example 5.2.9. $(E_{11}, E_{21} \dots, E_{m1}, E_{12}, \dots, E_{m2}, \dots, E_{1n}, \dots, E_{mn})$ is **linearly independent** in $M_{m\times n}(\mathbb{R})$.

Example 5.2.10. $(1, x, x^2, \dots, x^n)$ is **linearly independent** in $\mathbb{R}_n[x]$. In fact, the set $\{1\} \cup \{x^k \mid k \in \mathbb{N}\}$ is **linearly independent** in $\mathbb{R}[x]$.

Example 5.2.11. Let $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3$ be the three polynomials in $\mathbb{R}_2[x]$ introduced in **Example** (5.2.7). Then $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ is **linearly independent**.

To see this, suppose that $\boldsymbol{f} = a\boldsymbol{f}_1 + b\boldsymbol{f}_2 + c\boldsymbol{f}_3 = 0$ (the zero polynomial). Then for every x, $\boldsymbol{f}(x) = 0$, in particular, for $x = 1, 2, 3$. Recall, that $\boldsymbol{f}_1(1) = \boldsymbol{f}_2(2) = \boldsymbol{f}_3(3) = 1$ and $\boldsymbol{f}_1(2) = \boldsymbol{f}_1(3) = \boldsymbol{f}_2(1) = \boldsymbol{f}_2(3) = \boldsymbol{f}_3(1) = \boldsymbol{f}_3(2) = 0$. From this it follows that $\boldsymbol{f}(1) = a$, $\boldsymbol{f}(2) = b$, $\boldsymbol{f}(3) = c$ and consequently, $a = b = c = 0$.

Example 5.2.12. Determine if the sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ from $M_{2\times 2}(\mathbb{R})$ is **linearly independent** where

$$\boldsymbol{m}_1 = \begin{pmatrix}1 & 2\\ 2 & 3\end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix}2 & 1\\ 3 & 2\end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix}1 & 4\\ 3 & 6\end{pmatrix}, \boldsymbol{m}_4 = \begin{pmatrix}3 & 4\\ 2 & 3\end{pmatrix}.$$

We have to determine if there are scalars c_1, c_2, c_3, c_4, not all zero, such that

$$c_1\boldsymbol{m}_1 + c_2\boldsymbol{m}_2 + c_3\boldsymbol{m}_3 + c_4\boldsymbol{m}_4 =$$

$$c_1\begin{pmatrix}1 & 2\\ 2 & 3\end{pmatrix} + c_2\begin{pmatrix}2 & 1\\ 3 & 2\end{pmatrix} + c_3\begin{pmatrix}1 & 4\\ 3 & 6\end{pmatrix} + c_4\begin{pmatrix}3 & 4\\ 2 & 3\end{pmatrix} = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}. \tag{5.22}$$

After multiplying the matrix $\boldsymbol{m}_i$ by the scalar c_i and adding up Equation (5.22) becomes

$$\begin{pmatrix}c_1 + 2c_2 + c_3 + 3c_4 & 2c_1 + c_2 + 4c_3 + 4c_4\\ 2c_1 + 3c_2 + 3c_3 + 2c_4 & 3c_1 + 2c_2 + 6c_3 + 3c_4\end{pmatrix} = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix} \tag{5.23}$$

In order for the latter to occur each entry of the matrix on the left hand side of (5.23) must be zero. This gives rise to the **homogeneous linear system**

$$\begin{array}{ccccc} c_1 & +2c_2 & +c_3 & +3c_4 & =0 \\ 2c_1 & +3c_2 & +3c_3 & +2c_4 & =0 \\ 2c_1 & +c_2 & +4c_3 & +4c_4 & =0 \\ 3c_1 & +2c_2 & +6c_3 & +3c_4 & =0 \end{array} \tag{5.24}$$

which has **coefficient matrix**

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 2 & 3 & 3 & 2 \\ 2 & 1 & 4 & 4 \\ 3 & 2 & 6 & 3 \end{pmatrix} \tag{5.25}$$

Note that if in general we associate a 4-vector with an element of $M_{2\times 2}(\mathbb{R})$ by taking a matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ to $\begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{pmatrix}$ then the columns of the matrix in (5.25) are the vectors associated with $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4$. Moreover, the sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is **linearly independent** if and only if the sequence of 4-vectors

$$\left(\begin{pmatrix} 1 \\ 2 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 4 \\ 6 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 4 \\ 3 \end{pmatrix} \right)$$

is **linearly independent** in $\mathbb{R}^4$. In turn, the sequence in $\mathbb{R}^4$ is **linearly independent** if and only if the each of the columns of the matrix (5.25) is a **pivot column**. Finally, each of the columns of the matrix (5.25) is a **pivot column** if and only if the **null space** of the matrix (5.25) is the **zero subspace** of $\mathbb{R}^4$.

The **reduced echelon form** of the matrix (5.25) is $\begin{pmatrix} 1 & 0 & 0 & 25 \\ 0 & 1 & 0 & -6 \\ 0 & 1 & 0 & -10 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and so we conclude that $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3 \boldsymbol{m}_4)$ is **linearly dependent**.

We return to the development of our theory with a general result on **linearly dependent sequences**.

Theorem 5.2.5. *Let S be a sequence of vectors from a* **vector space** *V. Then the following hold:*

1. If S consists of a single vector $\boldsymbol{v}$ then S is **linearly dependent** *if and only if $\boldsymbol{v} = \mathbf{0}$.*

2. If $S = (\boldsymbol{u}, \boldsymbol{v})$ and both vectors are non-zero, then S is **linearly dependent** *if and only if the vectors are scalar multiples of one another.*

3. If $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ with $k \geq 2$ then S is **linearly dependent** *if and only if at least one of the vectors in a* **linear combination** *of the remaining vectors.*

4. Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ with $k \geq 2$ and assume for some i with $1 \leq i < k$ that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i)$ is **linearly independent**. *Then S is* **linearly dependent** *if and only if there is a $j > i$ such that $\boldsymbol{v}_j$ is a* **linear combination** *of the sequence of vectors which proceed it: $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j-1})$.*

The proof of 1) is the same as **Theorem** (2.4.5). The proof of 2) mimics that of **Theorem** (2.4.6). Finally, 3) and 4) can be proved in exactly the same way as 1) and 2) of **Theorem** (2.4.7) from Section (2.4).

We also state a result the proof of which is exactly like that of **Theorem** (2.4.8):

Theorem 5.2.6.

1. Any sequence or set of vectors which contains the **zero vector** *is* **linearly dependent**.

2. Any sequence which contains repeated vectors is **linearly dependent**.

3. Any set which contains a **linearly dependent** *subset is itself linearly dependent.*

4. Any subset of a **linearly independent** *set of vectors is linearly independent.*

We state one further result which generalizes a theorem from Section (2.4) to abstract **vector spaces**.

Theorem 5.2.7. *Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be a* **linearly independent** *sequence of vectors in a* **vector space** *V. Then the following hold:*

1. Any vector $\boldsymbol{v}$ in the **span** *of S is expressible in one and only one way as a* **linear combination** *of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.*

2. If $\boldsymbol{v}$ is not in the **span** *of S then we get a* **linearly independent** *sequence by adjoining $\boldsymbol{v}$ to S, that is, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k, \boldsymbol{v})$ is linearly independent.*

Remark 5.4. Theorem (5.2.7) suggests the virtue of having a sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ which is **linearly independent** and is a **spanning sequence** of V. This concept is so important we give it a name.

Definition 5.13. Let V be a nonzero **vector space**. A sequence $\mathcal{B}$ of V is said to be a ***basis*** of V if

1. $\mathcal{B}$ is **linearly independent**; and

2. $Span(\mathcal{B}) = V$.

More generally, a subset $\mathcal{B}$ of a **vector space** V is a ***basis*** of V if it is **linearly independent** and a **spanning set**.

Remark 5.5. The second part of the definition is for spaces like $\mathbb{R}[x]$. As remarked earlier, this space does not have any **spanning sequences** but does have spanning sets. In the absence of allowing infinite spanning sets such a space could not possibly have a basis.

The next theorem is beyond the sophistication of this book and will not be proved. It depends on a result from **set theory** called **Zorn's lemma** which is equivalent to one of the fundamental assumptions known as **axiom of choice**. In proving the theorem in its full generality the difficulty lies with the possiblity of infinite bases. When this possibility is eliminated the theorem becomes accessible and we prove it in the next section.

Theorem 5.2.8. *Every* **vector space** *has a* **basis***.*

Example 5.2.13. $(1, x, \ldots, x^n)$ is a **basis** for $\mathbb{R}_n[x]$. This is referred to as the ***standard basis*** of $\mathbb{R}_n[x]$.

Example 5.2.14. Recall, for $1 \leq i \leq m$ and $1 \leq j \leq n$ that E_{ij} is the $m \times n$ matrix which has all entries equal to zero except the $(i, j)-$entry, which is one. Then $(E_{11}, \ldots, E_{m1}, E_{12}, \ldots, E_{m2}, \ldots, E_{1n}, \ldots, E_{mn})$ is a basis for $M_{m\times n}(\mathbb{R})$. This is referred to as the ***standard basis of*** $M_{m\times n}(\mathbb{R})$

Example 5.2.15. The sequence $(\boldsymbol{f}_1(x), \boldsymbol{f}_2(x), \boldsymbol{f}_3(x))$ from **Example** (5.2.7) is a **basis** for $\mathbb{R}_2[x]$.

More generally, let $\boldsymbol{g}_1(x), \boldsymbol{g}_2(x), \boldsymbol{g}_3(x)$ be three polynomials in $\mathbb{R}_2[x]$. Let A be the matrix whose $(i,j)-$entry is $\boldsymbol{g}_i(j)$. Then $(\boldsymbol{g}_1(x), \boldsymbol{g}_2(x), \boldsymbol{g}_3(x))$ is a basis of $\mathbb{R}_2[x]$ if and only if A is invertible. Proving this is left as a **challenge exercise**.

Example 5.2.16. $\mathcal{B} = (1-x^3, x-x^3, x^2-x^3, 1+x+x^2+x^3)$ is a basis for $\mathbb{R}_3[x]$.

We can prove that this sequence is **linearly independent** and a **spanning sequence** of $\mathbb{R}_3[x]$ all at once. Suppose we want to write $A+Bx+Cx^2+Dx^3$ as a **linear combination** of $\mathcal{B}$. Then we need to show that there are scalars a, b, c, d such that

$$a(1-x^3)+b(x-x^3)+c(x^2-x^3)+d(1+x+x^2+x^3) = A+Bx+Cx^2+Dx^3.$$

We collect terms on the left hand side:

$$(a+d)+(b+d)x+(c+d)x^2+(-a-b-c+d)x^3 = A+Bx+Cx^2+Dx^3.$$

This leads to the **linear system**

$$\begin{array}{rcrcrcrcl} a & + & & & & + & d & = & A \\ & & b & & & + & d & = & B \\ & & & & c & + & d & = & C \\ -a & - & b & - & c & + & d & = & D \end{array}$$

This **linear system** has the **augmented matrix**

$$\begin{pmatrix} 1 & 0 & 0 & 1 & | & A \\ 0 & 1 & 0 & 1 & | & B \\ 0 & 0 & 1 & 1 & | & C \\ -1 & -1 & -1 & 1 & | & D \end{pmatrix}.$$

Note once again that the columns of this matrix are the coefficients of $1, x, x^2, x^3$ of the vectors in the sequence $\mathcal{B}$. We use **Gaussian elimination** to obtain an **echelon form**:

$$\to \begin{pmatrix} 1 & 0 & 0 & 1 & | & A \\ 0 & 1 & 0 & 1 & | & B \\ 0 & 0 & 1 & 1 & | & C \\ 0 & 0 & 0 & 4 & | & A+B+C+D \end{pmatrix}.$$

At this point we see that every column is a **pivot column** and therefore if $A = B = C = D = 0$ it then follows that the only solution is $a = b = c = d = 0$. This means that the sequence $\mathcal{B}$ is **linearly independent**. But we get more: for every A, B, C, D there is a unique **solution** which implies that $\mathcal{B}$ is also a **spanning sequence** and consequently a **basis** of $\mathbb{R}_3[x]$.

We conclude with an important result characterizing those sequences of vectors $\mathcal{B}$ which are **bases** of a **vector space** V.

Theorem 5.2.9. *A sequence $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of a **vector space** V is a **basis** of V if and only if for each vector $\boldsymbol{v}$ in V there is a unique sequence of scalars $(c_1, c_2, \dots, c_k)$ such that $\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$.*

Proof. Suppose $\mathcal{B}$ is a basis and $\boldsymbol{u}$ is a vector in V. Since $\mathcal{B}$ is a **spanning sequence** of V there are scalars $c_1, c_2, \dots, c_k$ such that $\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$. This shows that every vector is in $Span(\mathcal{B})$. By **Theorem** (5.2.7), every vector $\boldsymbol{u}$ which is in $Span(\mathcal{B}) = V$ has a unique representation as a **linear combination** of $\mathcal{B}$.

On the other hand assume every vector $\boldsymbol{u}$ is uniquely representable as a **linear combination** of $\mathcal{B}$. Since every vector is representable, this implies $Span(\mathcal{B}) = V$. On the other hand, the uniqueness assumption applies, in particular, to the zero vector $\mathbf{0}_V$. However, $\mathbf{0}_V = 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_k$. Since there is a unique representation for every vector $\boldsymbol{u}$ there is only the **trivial dependence relation** and consequently $\mathcal{B}$ is also **linearly independent**. $\square$

What You Can Now Do

1. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V and a vector $\boldsymbol{u}$ in V, determine if $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$, that is, if $\boldsymbol{u}$ is in **span** of $(\boldsymbol{v}_1, \dots \boldsymbol{v}_k)$. If $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ write $\boldsymbol{u}$ as a **linear combination** of $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$.

2. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V, determine if $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k) = V$.

3. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Determine if a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V is **linearly dependent** or **linearly independent**.

4. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V is **linearly dependent**, find a **non-trivial dependence relation**.

5. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V is **linearly dependent**, express one of the vectors as a **linear combination** of the remaining vectors.

6. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Determine if a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V is a **basis** of V.

Methods (How To Do It)

Method 5.2.1. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V and a vector $\boldsymbol{u}$ in V, determine if $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$.

If V is $\mathbb{R}_n[x]$ associate to each vector $\boldsymbol{v} = a_0 + a_1x + \cdots + a_nx^n$ in V the (n+1)-vector $\boldsymbol{v}' = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}$.

As an example, when $n = 3$ we associate to the polynomial $1 - 3x + 5x^2 - 7x^3$ the 4-vector $\begin{pmatrix} 1 \\ -3 \\ 5 \\ -7 \end{pmatrix}$. On the other hand, to the polynomial $-2x + 8x^2$ we associate the 4-vector $\begin{pmatrix} 0 \\ -2 \\ 8 \\ 0 \end{pmatrix}$.

If V is $M_{m\times n}(\mathbb{R})$ associate to each vector $\boldsymbol{v} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$ the (mn)-vector

$$\boldsymbol{v}' = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \\ a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \\ \vdots \\ a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

For example if $(m, n) = (3, 2)$ then we associate to the matrix $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ the 6-vector $\begin{pmatrix} 1 \\ 4 \\ 2 \\ 5 \\ 3 \\ 6 \end{pmatrix}$. On the other hand we associate to the matrix $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 3 & 0 \end{pmatrix}$ the 6-vector $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 3 \\ -1 \\ 0 \end{pmatrix}$.

In either case form the matrix $A = (\boldsymbol{v}_1' \ \boldsymbol{v}_2' \ \dots \ \boldsymbol{v}_k' \mid \boldsymbol{u}')$. Use **Method** (1.2.4) to determine if the **linear system** with **augmented matrix** equal to A is **consistent**: apply **Gaussian elimination** to obtain an **echelon form**. If the last column is not a **pivot column** then $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$, otherwise it is not. If $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ continue with **Gaussian elimination** to obtain the **reduced echelon form**. Find a **solution** $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ to the **linear system**. Then $\boldsymbol{u} = c_1\boldsymbol{v}_1 + \dots + c_k\boldsymbol{v}_k$.

Remark 5.6. **Method** (5.2.1) is suggested by **Example** (5.2.1), **Example** (5.2.2), and **Example** (5.2.3). It will become clearer (and rigorously demonstrated) why this works in a subsequent section.

Example 5.2.17. Let $V = M_{2\times 2}(\mathbb{R})$, and

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 1 \\ 1 & -3 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 1 & 2 \\ 2 & -5 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 3 & 1 \\ 2 & -6 \end{pmatrix}.$$

Determine if $\boldsymbol{m} = \begin{pmatrix} 4 & -2 \\ 1 & -2 \end{pmatrix}$ is in the **span** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$. If so, then express $\boldsymbol{m}$ as a **linear combination** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$.

We make the augmented matrix

$$\begin{pmatrix} 1 & 1 & 3 & | & 4 \\ 1 & 2 & 2 & | & 1 \\ 1 & 2 & 1 & | & -2 \\ -3 & -5 & -6 & | & -2 \end{pmatrix}.$$

We use **Gaussian elimination** to get the **echelon form**:

$$\to \begin{pmatrix} 1 & 1 & 3 & | & 4 \\ 0 & 1 & -1 & | & -3 \\ 0 & 1 & -2 & | & -6 \\ 0 & -2 & 3 & | & 10 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 3 & | & 4 \\ 0 & 1 & -1 & | & -3 \\ 0 & 0 & -1 & | & -3 \\ 0 & 0 & 1 & | & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 3 & | & 4 \\ 0 & 1 & -1 & | & -3 \\ 0 & 0 & -1 & | & -3 \\ 0 & 0 & 0 & | & 1 \end{pmatrix}$$

The last column is a **pivot column** and therefore $\boldsymbol{m}$ is not a **linear combination** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$, or, what amounts to the same thing, $\boldsymbol{m} \notin Span(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$.

Example 5.2.18. We continue with the notation of the previous example. Now determine if $\boldsymbol{m}' = \begin{pmatrix} 1 & -5 \\ 2 & 2 \end{pmatrix}$ is in the **span** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$. If $\boldsymbol{m}' \in Span(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$ then express $\boldsymbol{m}'$ as a **linear combination** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$.

Now we make the augmented matrix

$$\begin{pmatrix} 1 & 1 & 3 & | & 1 \\ 1 & 2 & 2 & | & 3 \\ 1 & 2 & 1 & | & 4 \\ -3 & -5 & -6 & | & -8 \end{pmatrix}.$$

We use **Gaussian elimination** to get the **echelon form**:

$$\to \begin{pmatrix} 1 & 1 & 3 & | & 1 \\ 0 & 1 & -1 & | & 2 \\ 0 & 1 & -2 & | & 3 \\ 0 & -2 & 3 & | & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 3 & | & 1 \\ 0 & 1 & -1 & | & 2 \\ 0 & 0 & -1 & | & 1 \\ 0 & 0 & 1 & | & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 3 & | & 1 \\ 0 & 1 & -1 & | & 2 \\ 0 & 0 & -1 & | & 1 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

At this point we see that the answer is yes, $\boldsymbol{m}'$ is in the **span** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$.

We continue with the **Gaussian elimination** to obtain the **reduced echelon form**:

$$\to \begin{pmatrix} 1 & 1 & 3 & | & 1 \\ 0 & 1 & -1 & | & 2 \\ 0 & 0 & 1 & | & -1 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 & | & 4 \\ 0 & 1 & 0 & | & 1 \\ 0 & 0 & 1 & | & -1 \\ 0 & 0 & 0 & | & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & | & 3 \\ 0 & 1 & 0 & | & 1 \\ 0 & 0 & 1 & | & -1 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

There is a unique **solution** to this **linear system** consisting of the single vector $\begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}$.

Thus, $\boldsymbol{m}' = 3\boldsymbol{m}_1 + \boldsymbol{m}_2 - \boldsymbol{m}_3$.

Example 5.2.19. Let $\boldsymbol{f}_1 = 1+x+x^2+x^4$, $\boldsymbol{f}_2 = 2+3x-x^2+2x^3+2x^4$, $\boldsymbol{f}_3 = 2+x+4x^2-x^3+2x^4 \in \mathbb{R}_4[x]$. Determine if the polynomial $\boldsymbol{f} = x^2+x^3 \in Span(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ and, if so, express $\boldsymbol{f}$ as a **linear combination** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$.

We associate these polynomials to 5-vectors. The vectors we obtain are:

$$\boldsymbol{f}_1 \to \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{f}_2 \to \begin{pmatrix} 2 \\ 3 \\ -1 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{f}_3 \to \begin{pmatrix} 2 \\ 1 \\ 4 \\ -1 \\ 2 \end{pmatrix}, \boldsymbol{f} \to \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

We make the augmented matrix

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

We use **Gaussian elimination** to obtain an **echelon form**:

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

So, we see that $\boldsymbol{f} \notin Span(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$, that is, $\boldsymbol{f}$ is not in the **span** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$.

Example 5.2.20. Continuing with the previous example, now determine if $\boldsymbol{f}' = x^2 - x^3$ is n the **span** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$. If $\boldsymbol{f}' \in Span(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ then express $\boldsymbol{f}'$ as a **linear combination** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$.

Now $\boldsymbol{f}'$ is associated to the 5-vector $\begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$. The augmented matrix we obtain is

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

We use **Gaussian elimination** to obtain an **echelon form**:

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

In this case $\boldsymbol{f}'$ is a in the **span** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$. We now find scalars c_1, c_2, c_3 such that $\boldsymbol{f}' = c_1\boldsymbol{f}_1 + c_2\boldsymbol{f}_2 + c_3\boldsymbol{f}_3$ by a continuing with **Gaussian elimination** to obtain the **reduced echelon form** of the matrix"

$$\rightarrow \left(\begin{array}{ccc|c} 1 & 2 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 2 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The **linear system** has a unique solution, $\begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}$, and therefore $\boldsymbol{f}' = 4\boldsymbol{f}_1 - \boldsymbol{f}_2 - \boldsymbol{f}_3$.

Method 5.2.2. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Given a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of vectors in V, determine if $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k) = V$.

As in **Method** (5.2.1) we associate with each of the vectors $\boldsymbol{v}_i$ an l-vector $\boldsymbol{v}_i'$ where $l = n+1$ if $V = \mathbb{R}_n[x]$ and $l = mn$ if $V = M_{m\times n}(\mathbb{R})$. Then we form $(\boldsymbol{v}_1' \ \ldots \ \boldsymbol{v}_k')$, the matrix with the vectors $\boldsymbol{v}_i'$ as columns and use **Gaussian elimination** to obtain an **echelon form**. If every row has a **pivot position** then the **span** of $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is V and otherwise it is not V (see also **Method** (2.3.2)).

Example 5.2.21. Let $\boldsymbol{m}_1 = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 1 & 3 \\ -2 & -2 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 3 & 2 \\ -2 & -3 \end{pmatrix}$ be vectors from $M_{2\times 2}(\mathbb{R})$. Determine if $Span(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3) = M_{2\times 2}(\mathbb{R})$, that is, the **span** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$ is $M_{2\times 2}(\mathbb{R})$.

This is fairly easy and we will see that no computation is necessary: the corresponding columns are:

$$\boldsymbol{m}_1 \to \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{m}_2 \to \begin{pmatrix} 1 \\ -2 \\ 3 \\ -2 \end{pmatrix}, \boldsymbol{m}_3 \to \begin{pmatrix} 3 \\ -2 \\ 2 \\ -3 \end{pmatrix}.$$

The matrix we obtain is

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

Since this matrix is 4×3 it is not possible for every row to contain a **pivot position**. In fact, $M_{2\times 2}(\mathbb{R})$ cannot be spanned with fewer than 4 vectors and, in general, $M_{m\times n}(\mathbb{R})$ cannot be spanned by fewer then mn vectors.

Example 5.2.22. Continue with the notation of the previous example. Suppose $\boldsymbol{m}_4 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ which corresponds to the column $\begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix}$. Determine if $Span(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4) = M_{2\times 2}(\mathbb{R})$.

Now the matrix is

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

We apply **Gaussian elimination** to obtain an **echelon form**:

$$\to \begin{pmatrix} 1 & 1 & 3 & 1 \\ 0 & -1 & 1 & 4 \\ 0 & 2 & -1 & 1 \\ 0 & -1 & 0 & 5 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 3 & 1 \\ 0 & -1 & 1 & 4 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & -1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 3 & 1 \\ 0 & -1 & 1 & 4 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}.$$

Thus, we see that the **span** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is $M_{2\times 2}(\mathbb{R})$ since every row has a **pivot position**.

Method 5.2.3. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Determine if a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of vectors in V is **linearly dependent** or **linearly independent**.

Again, we turn the vectors into columns and make a matrix. Then we determine if the sequence of columns is **linearly independent**. Recall, this is done by **Method** (2.4.1): We apply **Gaussian elimination** to obtain an **echelon form**. If every column is a **pivot column** then the sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ from V is **linearly independent**, otherwise it is **linearly dependent**.

Example 5.2.23. Let $\boldsymbol{f}_1 = 1 + x + x^2, \boldsymbol{f}_2 = 2 + x + 2x^2, \boldsymbol{f}_3 = 1 + 2x + 2x^2, \boldsymbol{f}_4 = 1 + 2x + 3x^2$.

These correspond to the following columns:

$$\boldsymbol{f}_1 \to \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{f}_2 \to \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \boldsymbol{f}_3 \to \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{f}_4 \to \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

We form the matrix

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

Since this matrix is 3×4 it has at most three **pivot columns** so it is not possible for every column to be a pivot column and we can immediately conclude that the sequence $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$ is necessarily **linearly dependent**. In fact, any sequence of four vectors in $\mathbb{R}_2[x]$ is **linearly dependent** and, more generally, any sequence of $n+2$ vectors in $\mathbb{R}_n[x]$ is linearly dependent. A similar argument implies that any set of $mn + 1$ vectors in $M_{m\times n}(\mathbb{R})$ is **linearly dependent**.

Let's determine if the sequence $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ is **linearly dependent** or **linearly independent**.

Now the matrix is

$$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix}.$$

We apply **Gaussian elimination** to obtain an **echelon form**:

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

Every column is a **pivot column** and so $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ is **linearly independent**.

Method 5.2.4. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ of vectors in V is **linearly dependent**, find a **non-trivial dependence relation**.

Associate to each of the vectors $\boldsymbol{v}_i$ an $l-$vector, $\boldsymbol{v}_i'$ where $l = n+1$ if $V = \mathbb{R}_n[x]$ and $l = mn$ if $V = M_{m\times n}(\mathbb{R})$ as previously described in **Method** (5.2.1). Then form the $l \times k$ matrix $A = (\boldsymbol{v}_1' \ \dots \ \boldsymbol{v}_k')$ with these as its columns. Use **Method** (3.2.2) to find a non-zero $k-$vector in the **null space** of A. This gives the coefficients of a **non-trivial dependence relation**.

Example 5.2.24. We previously demonstrated in **Example** (5.2.23) that the sequence of vectors $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$ from $\mathbb{R}_2[x]$ where $\boldsymbol{f}_1 = 1+x+x^2, \boldsymbol{f}_2 = 2+x+2x^2, \boldsymbol{f}_3 = 1+2x+2x^2, \boldsymbol{f}_4 = 1+2x+3x^2$ is **linearly dependent**. The matrix we obtain is

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

We apply **Gaussian elimination** to obtain the **reduced echelon form**. We then write out the **homogeneous linear system** which has this **echelon form** as its **coefficient matrix**. We separate the **free and leading variables**, set the free variables equal to parameters and solve for each variable in terms of these parameters.

We proceed to do apply **Gaussian elimination** :

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

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

The **homogeneous linear system** with this matrix is

$$\begin{array}{ccccccc} x & & & - & 3w & = & 0 \\ & y & & + & w & = & 0. \\ & & z & + & 2w & = & 0 \end{array}$$

x, y, z are the **leading variables** and there is one **free variable**, w. Setting $w = t$, we conclude that the **solution set** to the **homogeneous linear system** is $\{\begin{pmatrix} 3t \\ -t \\ -2t \\ t \end{pmatrix} | t \in \mathbb{R}\}$ which is the **span** of $\begin{pmatrix} 3 \\ -1 \\ -2 \\ 1 \end{pmatrix}$. Therefore $3\boldsymbol{f}_1 - \boldsymbol{f}_2 - 2\boldsymbol{f}_3 + \boldsymbol{f}_4 = \mathbf{0}_{\mathbb{R}_2[x]}$.

Method 5.2.5. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. If a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of vectors in V is **linearly dependent**, express one of the vectors as a **linear combination** of the remaining vectors.

Use **Method** (5.2.4) to obtain a **non-trivial dependence relation** for the sequence. Then any vector in the sequence that does not have a zero coefficient can be solved in terms of the remaining vectors.

Example 5.2.25. In **Example** (5.2.23) we demonstrated that the sequence of vectors $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$ from $\mathbb{R}_2[x]$ where $\boldsymbol{f}_1 = 1 + x + x^2, \boldsymbol{f}_2 = 2 + x + 2x^2, \boldsymbol{f}_3 = 1 + 2x + 2x^2, \boldsymbol{f}_4 = 1 + 2x + 3x^2$ is **linearly dependent** In **Example** (5.2.24) we found the **non-trivial dependence relation**

$$3\boldsymbol{f}_1 - \boldsymbol{f}_2 - 2\boldsymbol{f}_3 + \boldsymbol{f}_4 = \mathbf{0}_{\mathbb{R}_2[x]}.$$

Since each coefficient is non-zero each vector can be solved for in terms of the remaining vectors:

$$\boldsymbol{f}_1 = \frac{1}{3}\boldsymbol{f}_2 + \frac{2}{3}\boldsymbol{f}_3 - \frac{1}{3}\boldsymbol{f}_4$$

$$\boldsymbol{f}_2 = 3\boldsymbol{f}_1 - 2\boldsymbol{f}_3 + \boldsymbol{f}_4$$

$$\boldsymbol{f}_3 = \frac{3}{2}\boldsymbol{f}_1 - \frac{1}{2}\boldsymbol{f}_2 + \frac{1}{2}\boldsymbol{f}_4$$

$$\boldsymbol{f}_4 = -3\boldsymbol{f}_1 + \boldsymbol{f}_2 + 2\boldsymbol{f}_3$$

Method 5.2.6. Assume V is the space $\mathbb{R}_n[x]$ or $M_{m\times n}(\mathbb{R})$. Determine if a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of vectors in V is a **basis** of V.

We continue to turn the vectors into columns and make a matrix. We use **Gaussian elimination** to get an **echelon form**. The sequence is a **basis** if and only if every row has a **pivot position** and every column is a **pivot column**. In turn, this occurs if and only if the matrix is **square** and **invertible**.

Example 5.2.26. The sequence of vectors $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ introduced in **Example** (5.2.23) is a basis of $\mathbb{R}_2[x]$.

Example 5.2.27. Let $\boldsymbol{g}_1 = 1-x, \boldsymbol{g}_2 = x-x^2, \boldsymbol{g}_3 = x^2-x^3, \boldsymbol{g}_4 = 1+x+x^2-3x^3$. Determine if $(\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3, \boldsymbol{g}_4)$ is a **basis** of $\mathbb{R}_3[x]$.

The matrix we obtain is

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

Applying **Gaussian elimination** we obtain the **echelon form**

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

So, the sequence is not a **basis** of $\mathbb{R}_3[x]$.

Example 5.2.28. In the previous example, do we obtain a basis if we replace $\boldsymbol{g}_4$ by $\boldsymbol{g}_4' = 1 + x + x^2 + x^3$?

Now the matrix is

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

An **echelon form** of this matrix, obtained by **Gaussian elimination**, is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

Therefore $(\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3, \boldsymbol{g}_4')$ is a basis of $\mathbb{R}_3[x]$.

Example 5.2.29. Let $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3$ be the vectors from $M_{2\times 2}(\mathbb{R})$ of **Example** (5.2.17). Let $\boldsymbol{m}_4 = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$. Determine if $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$.

The matrix we obtain is

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

Using **Gaussian elimination** we obtain the following **echelon form**

$$\begin{pmatrix} 1 & 1 & 3 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 10 \end{pmatrix}.$$

Since this is a **square matrix** and every column is a **pivot column**, the sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$.

Exercises

Throughout these exercises

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix}, \boldsymbol{m}_4 = \begin{pmatrix} 0 & 2 \\ 9 & -3 \end{pmatrix},$$

$$\boldsymbol{m}_5 = \begin{pmatrix} 2 & 2 \\ 6 & 3 \end{pmatrix}, \boldsymbol{m}_6 = \begin{pmatrix} 3 & 4 \\ 7 & 4 \end{pmatrix}, \boldsymbol{m}_7 = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}, \boldsymbol{m}_8 = \begin{pmatrix} 0 & 2 \\ 0 & -1 \end{pmatrix}, \boldsymbol{m}_9 = \begin{pmatrix} 1 & 0 \\ 5 & 4 \end{pmatrix}$$

$$\boldsymbol{f}_1 = 1 + x + x^2 + x^4, \boldsymbol{f}_2 = x + x^2 + x^3 + x^4, \boldsymbol{f}_3 = x + 2x^2 + x^4,$$

$$\boldsymbol{f}_4 = 1 + 3x + 4x^2 + 2x^3, \boldsymbol{f}_5 = 1 - x^4, \boldsymbol{f}_6 = 1 - x^3, \boldsymbol{f}_7 = 1 + x^4,$$

$$\boldsymbol{f}_8 = 1 - x^2, \boldsymbol{f}_9 = 1 + 2x^3 - 2x^4, \boldsymbol{f}_{10} = x + 2x^2 + 3x^3 + 4x^4.$$

For 1-4 see **Method** (5.2.1).

1. Determine if $\boldsymbol{m}_4$ is in the **span** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$.

2. Determine if $\boldsymbol{m}_5$ is in the **span** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$.

3. Determine if $\boldsymbol{f}_5$ is in the **span** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$.

4. Determine if $\boldsymbol{f}_6$ is in the **span** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$.

For exercises 5, 6 see **Method** (5.2.2).

5. Determine if $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_6)$ is a **spanning sequence** of $M_{2\times 2}(\mathbb{R})$.

6. Determine if $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_7)$ is a **spanning sequence** of $\mathbb{R}_4[x]$.

For exercises 7-9 see **Method** (5.2.3).

7. Determine if $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_8)$ is **linearly independent or linearly dependent**.

8. Determine if $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_5)$ is **linearly independent or linearly dependent**.

9. Determine if $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_{10})$ is **linearly independent or linearly dependent**.

For exercises 10-13 see **Method** (5.2.6).

10. Determine if $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_8)$ is a **basis** of $\mathbb{R}_4[x]$.

11. Determine if $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_9)$ is a **basis** of $\mathbb{R}_4[x]$.

12. Determine if $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_7)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$.

13. Determine if $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_4, \boldsymbol{m}_9)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$.

14. Show that $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_9)$ is **linearly dependent** and exhibit one **non-trivial dependence relation**. Then express $\boldsymbol{m}_9$ as a **linear combination** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$. See **Method** (5.2.3), **Method** (5.2.4), and **Method** (5.2.5).

In exercises 15 - 20 answer true or false and give an explanation.

15. If the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ **spans** the space V and $\boldsymbol{u} \in V$ then the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k, \boldsymbol{u})$ is **linearly dependent**.

16. If the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ **spans** the space V then the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is a **basis** of V.

17. If the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ **spans** the space V then the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_{k-1})$ cannot span V.

18. If the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ **spans** the space V and $\boldsymbol{u}$ is a vector in V then the sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{u})$ spans V.

19. If the sequence of polynomials $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_5)$ from $\mathbb{R}_4[x]$ is **linearly independent** then the sequence $(\boldsymbol{f}_1, \boldsymbol{f}_3, \boldsymbol{f}_5)$ is s **linearly independent**.

20. The sequence $(1, x, x^2)$ is the unique **basis** of $\mathbb{R}_2[x]$.

Challenge Exercises (Problems)

For a polynomial $f(x) = c_0 + c_1x + c_2x + \cdots + c_dx^d$ and an $n \times n-$matrix A interpret $f(A)$ to be $c_0I_n + c_1A + c_2A^2 + \cdots + c_dA^d$ (here I_n is the $n \times n$ identity matrix.)

1. Let A be a $2 \times 2-$matrix. Prove there is some polynomial $f(x) \in \mathbb{R}_4[x]$ such that

$$f(A) = \mathbf{0}_{M_{2\times 2}(\mathbb{R})}.$$

Hint: Think about the sequence of vectors $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, A, A^2, A^3, A^4$ from $M_{2\times 2}(\mathbb{R})$.

2. Let A be a $3 \times 3-$matrix and $\boldsymbol{v} \in \mathbb{R}^3$. Prove that there is some polynomial $g(x) \in \mathbb{R}_3[x]$ such that

$$g(A)\boldsymbol{v} = \mathbf{0}_3.$$

3. a) Verify that the matrix

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

is **invertible**.

b) Let V be a **vector space** and $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ be an **linearly independent** sequence of vectors in V.

Set $\boldsymbol{w}_1 = \boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{w}_2 = \boldsymbol{v}_2 + \boldsymbol{v}_3, \boldsymbol{w}_3 = \boldsymbol{v}_1 - 2\boldsymbol{v}_3$. Prove that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is a **linearly independent** sequence of vectors in V.

4. Let $\boldsymbol{g}_1(x), \boldsymbol{g}_2(x), \boldsymbol{g}_3(x)$ be in $\mathbb{R}_2[x]$. Set $a_{ij} = \boldsymbol{g}_j(i)$. Prove that $(\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3)$ is a basis for $\mathbb{R}_2[x]$ if and only if the matrix (a_{ij}) is **invertible**.

5. Set

$$\boldsymbol{f}_1(x) = \frac{(x-2)(x-3)(x-4)}{-6}$$

$$\boldsymbol{f}_2 = \frac{(x-1)(x-3)(x-4)}{2}$$

$$\boldsymbol{f}_3 = \frac{(x-1)(x-2)(x-4)}{-2}$$

$$\boldsymbol{f}_4 = \frac{(x-1)(x-2)(x-3)}{6}$$

Prove that $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$ is a **basis** of $\mathbb{R}_3[x]$.

Quiz Solutions

1. The **reduced echelon form** of the matrix $\begin{pmatrix} -1 & 5 & 1 \\ 2 & 2 & 1 \\ 5 & -1 & 1 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 & \frac{1}{4} & \\ 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \end{pmatrix}$. Since the last column is not a **pivot column** the sequence is **linearly dependent**. Moreover,

$$(1)\begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix} + (1)\begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} + (-4)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Not right, see **Method** (2.4.1) and **Method** (2.4.2).

2. The **reduced echelon form** of the matrix $\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ -1 & 2 & 2 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Since every column is a **pivot column** the sequence **linearly independent**.

Not right, see **Method** (2.4.1).

3. We can conclude that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = \mathbb{R}^3$ from the **half is good enough theorem.**

4. Since $\boldsymbol{v}_1' = \begin{pmatrix} \boldsymbol{v}_1 \\ 0 \end{pmatrix}, \boldsymbol{v}_2' = \begin{pmatrix} \boldsymbol{v}_2 \\ -4 \end{pmatrix}, \boldsymbol{v}_3' = \begin{pmatrix} \boldsymbol{v}_3 \\ -5 \end{pmatrix}$. If $c_1\boldsymbol{v}_1' + c_2\boldsymbol{v}_2' + c_3\boldsymbol{v}_3' = \boldsymbol{0}_4$ then $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \boldsymbol{0}_3$. However, by 3) $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent** and consequently, $c_1 = c_2 = c_3 = 0$. Thus, $(\boldsymbol{v}_1', \boldsymbol{v}_2', \boldsymbol{v}_3')$ is **linearly independent**.

See **definition of linear independence**.

5. Assume $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$ are in S. Then

$$x_1 + x_2 + x_3 + x_4 = y_1 + y_2 + y_3 + y_4 = 0.$$

$\boldsymbol{x} + \boldsymbol{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ x_4 + y_4 \end{pmatrix}$. Then

$$(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) + (x_4 + y_4) =$$

$$(x_1 + x_2 + x_3 + x_4) + (y_1 + y_2 + y_3 + y_4) = 0 + 0 = 0$$

Thus, $\boldsymbol{x} + \boldsymbol{y}$ is in S and S is **closed under addition**.

Now assume $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ is in S so that

$$x_1 + x_2 + x_3 + x_4 = 0.$$

Assume c is a scalar. Then $c\boldsymbol{x} = \begin{pmatrix} cx_1 \\ cx_2 \\ cx_3 \\ cx_4 \end{pmatrix}$. Then

$$(cx_1) + (cx_2) + (cx_3) + (cx_4) =$$

$$c[x_1 + x_2 + x_3 + x_4] = c \times 0 = 0.$$

We can conclude that S is **closed under scalar multiplication** and a **subspace** of $\mathbb{R}^4$.

Not right, see the **definition of a subspace of $\mathbb{R}^n$**.

6. $(\boldsymbol{v}'_1, \boldsymbol{v}'_2, \boldsymbol{v}'_3)$ is a **linearly independent** sequence from S. We claim that $(\boldsymbol{v}'_1, \boldsymbol{v}'_2, \boldsymbol{v}'_3)$ is a **spanning sequence** of S. Suppose not, then there exists $\boldsymbol{u} \in S$ such that $\boldsymbol{u} \notin inSpan(\boldsymbol{v}'_1, \boldsymbol{v}'_2, \boldsymbol{v}'_3)$. However, it then follows by that the sequence $(\boldsymbol{v}'_1, \boldsymbol{v}'_2, \boldsymbol{v}'_3, \boldsymbol{u})$ is a **linearly independent** sequence in S. However, $(\boldsymbol{v}'_1, \boldsymbol{v}'_2, \boldsymbol{v}'_3, \boldsymbol{u})$ **spans** $\mathbb{R}^4$ by the **half is good enough theorem**. This contradicts the assumption that $(\boldsymbol{v}'_1, \boldsymbol{v}'_2, \boldsymbol{v}'_3, \boldsymbol{u})$ is a sequence from S, a proper **subspace** of $\mathbb{R}^4$.

5.3. Dimension of a finite generated vector space

We introduce the notion of a finite generated vector space (fgvs), prove that every fgvs has a basis, that all bases have the same number of elements, and define the notion of dimension.

Am I Ready for This Material

Extensive use is made of the following concepts and procedures in this section and they need to be mastered to fully understand the new ideas and methods introduced here:

vector space

subspace of a vector space

linear combination of a sequence of vectors

span of a sequence or set of vectors from a vector space

spanning sequence or set of a subspace of a vector space

linearly independent sequence or set of vectors from a vector space

linearly dependent sequence or set of vectors from a vector space

basis of a vector space

Quiz

1. Set

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\boldsymbol{m}_4 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \boldsymbol{m}_5 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Determine if the sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5)$ is **linearly independent**.

2. Let

$$\boldsymbol{f}_1 = 1 - 2x + x^3, \boldsymbol{f}_2 = 2 - 3x + x^2$$
$$\boldsymbol{f}_3 = 3 - 7x - x^2 + 5x^3, \boldsymbol{f}_4 = 2 + 3x - x^2 - 4x^3.$$

Determine if the sequence $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$ is **linearly independent**.

3. Set

$$m_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, m_2 = \begin{pmatrix} 2 & 1 \\ 3 & 3 \end{pmatrix}, m_3 = \begin{pmatrix} 2 & 5 \\ 1 & 4 \end{pmatrix}, m_4 = \begin{pmatrix} 1 & -3 \\ 4 & 1 \end{pmatrix}$$

Determine if the sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is a **spanning sequence** of $M_{2\times 2}(\mathbb{R})$.

4. Set

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 3 & -1 \\ 4 & 2 \end{pmatrix} \boldsymbol{m}_3 = \begin{pmatrix} 5 & -2 \\ 7 & 3 \end{pmatrix}, \boldsymbol{m}_4 = \begin{pmatrix} 3 & -1 \\ -1 & -1 \end{pmatrix}.$$

Determine if $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$.

5. Determine if the sequence $(1-x, 1-x^2, 1-x^3, 1+x^3, 1+x+x^2+x^3)$ is a **basis** of $\mathbb{R}_3[x]$.

Quiz Solutions

New Concepts

Three new concepts are introduced which are central to the study of **vector spaces**. These concepts are fundamental and will be used regularly from this point on. These are

finitely generated vector space

dimension of a finitely generated vector space

finite dimensional vector space

Theory (Why It Works)

In the previous section we stated the following result:

Theorem *Every* **vector space** *has a* **basis**.

As mentioned the proof of this is beyond the scope of this book but it is accessible for certain types of **vector spaces**, namely those which can be **spanned** by a finite sequence. We give a name to such vector spaces:

Definition 5.14. A **vector space** V is ***finitely generated*** if it is possible to find a finite sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ such that $V = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.

Example 5.3.1. The spaces $\mathbb{R}^n, \mathbb{R}_n[X], M_{m\times n}(\mathbb{R})$ are all **finitely generated**.

The spaces $\mathbb{R}[X], F(\mathbb{R}), C(\mathbb{R}), C^1(\mathbb{R})$ are not finitely generated.

Example 5.3.2. Let V be a **vector space** **spanned** by a sequence of length three, say $V = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. We claim the following: If S is a **linearly independent** sequence then the length of S is at most three.

Equivalently, if $S = (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$ then S is **linearly dependent**. Suppose to the contrary. We will get a contradiction.

The idea will be to exchange, one by one, elements of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ with elements of $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$ so that the new sequence is a **spanning sequence**.

Since $\boldsymbol{w}_4 \in V = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ we know that $(\boldsymbol{w}_4, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly dependent** by part 3) of **Theorem** (5.2.5).

We know that $\boldsymbol{w}_4$ is not zero since $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$ is **linearly independent**. By part 4) of **Theorem** (5.2.5) we can then conclude that one of the $\boldsymbol{v}_i, i = 1, 2, 3$, is a **linear combination** of the preceding vectors. By relabeling the $\boldsymbol{v}_i$ if necessary, we can assume that $\boldsymbol{v}_3$ is a **linear combination** of $(\boldsymbol{w}_4, \boldsymbol{v}_1, \boldsymbol{v}_2)$. By **Theorem** (5.2.4) we can remove $\boldsymbol{v}_3$ and the remaining sequence, $(\boldsymbol{w}_4, \boldsymbol{v}_1, \boldsymbol{v}_2)$ has the same **span** as $(\boldsymbol{w}_4, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ and therefore $Span(\boldsymbol{w}_4, \boldsymbol{v}_1, \boldsymbol{v}_2) = V$.

Now we can play the same game with $\boldsymbol{w}_3$ Since $\boldsymbol{w}_3 \in V = Span(\boldsymbol{w}_4, \boldsymbol{v}_1, \boldsymbol{v}_2)$ it follows by part 3) of **Theorem** (5.2.5) that $(\boldsymbol{w}_3, \boldsymbol{w}_4, \boldsymbol{v}_1, \boldsymbol{v}_2)$ is **linearly dependent**. Since $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$ is **linearly independent**, it follows that $(\boldsymbol{w}_3, \boldsymbol{w}_4)$ is linearly independent. It once again follows by part 4) of **Theorem** (5.2.5) that either $\boldsymbol{v}_1$ is a **linear combination** of $(\boldsymbol{w}_3, \boldsymbol{w}_4)$ or $\boldsymbol{v}_2$ is a linear combination of $(\boldsymbol{w}_3, \boldsymbol{w}_4, \boldsymbol{v}_1)$. By relabeling $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, if necessary, we can assume that $\boldsymbol{v}_2$ is a **linear combination** of $(\boldsymbol{w}_3, \boldsymbol{w}_4, \boldsymbol{v}_1)$.

Once again, by **Theorem** (5.2.4), we can remove $\boldsymbol{v}_2$ and conclude that $V = Span(\boldsymbol{w}_3, \boldsymbol{w}_4, \boldsymbol{v}_1)$. Repeating this argument one more time we get that $V = Span(\boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$. However, since $\boldsymbol{w}_1 \in V = Span(\boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$ it follows that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$ is **linearly dependent**, which is a contradiction.

In exactly the same way we can prove a more general result known as the **Exchange Theorem**:

Theorem 5.3.1. *Assume the* **vector space** *V can be generated by a sequence of n vectors. Than any sequence of of vectors of length $n+1$ (and hence greater) is* **linearly dependent**.

There are many wide ranging consequences of the exchange theorem. We come now to the first of these, the existence of a **basis** in a **finitely generated vector space**.

Theorem 5.3.2. *Let V be a* **finitely generated vector space***:*

$$V = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m).$$

Then V has a **basis** *of length at most m.*

Proof. By the **Exchange Theorem** no **linearly independent** sequence has more than m vectors. Choose a linearly independent sequence $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$ with n as large as possible. Such sequences exist because n must be less than or equal to m.

We claim that $Span(\mathcal{B}) = V$. We do a **proof by contradiction**: We assume that $Span(\mathcal{B}) \neq V$ and get a contradiction. If $V \neq Span(\mathcal{B})$ then there exists a vector $\boldsymbol{u}$ which is in V but not $Span(\mathcal{B})$. By part 2) of **Theorem** (5.2.7) it follows that the sequence $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n, \boldsymbol{u})$ of length $n+1$ is **linearly independent** which contradicts the maximality of n. □

We can use the same arguments to prove that every **spanning sequence** of a **finitely generated vector space** can be contracted to a **basis** and this will lead to an algorithm for finding a basis of a **vector space** given a **spanning sequence**.

The next result uses some of the same arguments to achieve a somewhat stronger statement:

Theorem 5.3.3. *Let V be a* **vector space** *and assume there is an integer n such that every* **linearly independent** *sequence has length at most n. Then V has a* **basis** *of length at most n.*

Because of the similarity to the above we leave the proof of this result as a **challenge exercise**.

Suppose now that V is a **finitely generated vector space** and has a **spanning sequence** of length m. Suppose W is a **subspace** of V. Then any **linearly independent** sequence of W is a linearly independent sequence of V and therefore has length at most m. Therefore **Theorem** (5.3.3) applies to W and we can conclude that W has a **basis**. We state this in the following:

Theorem 5.3.4. *Assume that V is a* **finitely generated vector space***. Then every* **subspace** *W of V has a* **basis***.*

A natural question arises: can there be bases with different lengths? The next theorem answers this question.

Theorem 5.3.5. *If a* **vector space** *V has a* **basis** *of length n then every basis of V has length n.*

Proof. Let $\mathcal{B}$ be a **basis** of length n and $\mathcal{B}'$ any other basis. Since $\mathcal{B}'$ is a **linearly independent** sequence and $\mathcal{B}$ is a **spanning sequence**, it follows by the **Exchange Theorem** that $\mathcal{B}'$ has length at most n. In particular, $\mathcal{B}'$ is finite. So let us suppose that $\mathcal{B}'$ has length m. We have just demonstrated that $m \leq n$.

On the other hand, since $\mathcal{B}'$ is a **basis** we have $Span(\mathcal{B}') = V$.

Because $\mathcal{B}$ is a **basis** it is **linearly independent**. Thus, by the **Exchange Theorem** $n \leq m$. Therefore we can conclude that $m = n$.

$\square$

Definition 5.15. Let V be a **finitely generated vector space**. The length n of a **basis** of V is called the ***dimension*** of V. We write $dim(V) = n$. Hereafter we will refer to a **finitely generated vector space** as a ***finite dimensional vector space***.

Example 5.3.3. 1. $dim(\mathbb{R}^n) = n$.

The sequence $(\boldsymbol{e}_1^n, \boldsymbol{e}_2^n, \ldots, \boldsymbol{e}_n^n)$ where $\boldsymbol{e}_i^n$ is a **basis** of $\mathbb{R}^n$. Here $\boldsymbol{e}_i^n$ is the **n-vector** all of whose components are 0 except the i^{th} component, which is 1.

2. $dim(\mathbb{R}_n[X]) = n + 1$.

The sequence $(1, x, x^2, \ldots, x^n)$ is a **basis**. This sequence has length $n + 1$.

3. $dim(M_{m\times n}(\mathbb{R})) = mn$.

Let E_{ij} be the matrix with all entries zero except in the $(i, j)-$entry, which is a 1. Then the sequence $(E_{11}, E_{21}, \ldots, E_{m1}, E_{12}, E_{22}, \ldots, E_{m2} \ldots, E_{1n}, E_{2n}, \ldots, E_{mn})$. is a **basis** of $M_{m\times n}(\mathbb{R})$. This sequence has length mn.

As a concrete example, the sequence

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

is a **basis** for $M_{3\times 2}(\mathbb{R})$.

Example 5.3.4. Let V be a **four dimensional vector space**.

1. Is it possible to **span** V with a sequence of length 3?

2. Can V have a **linearly independent** sequence of length 5?

The answer to both these is no as follows from the **Exchange Theorem**:

1. Let $\mathcal{B}$ be a **basis** of V. Suppose S is a **spanning sequence** of length 3. Then by the **Exchange Theorem** there are no **linearly independent** sequences of length 4. However, since $\mathcal{B}$ is a **basis**, in particular, $\mathcal{B}$ is linearly independent. Since $dim(V) = 4$, the length of $\mathcal{B}$ is 4 and we have a contradiction. Thus, there are no **spanning sequences** of V with length 3.

2. Suppose S is a **linearly independent** sequence of vectors. Since $\mathcal{B}$ is a **basis**, in particular, $\mathcal{B}$ is a **spanning sequence**. Since $dim(V) = 4$ the length of $\mathcal{B}$ is 4. It follows from the **Exchange Theorem** that the length of S is less than or equal 4.

These very same arguments are used to prove the following general theorem:

Theorem 5.3.6. *Let V be an **n-dimensional** **vector space**. Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$ be a sequence of vectors from V.*

*1. If S is **linearly independent** then $m \leq n$.*

*2. If S **spans** V then $m \geq n$.*

Suppose now that V is an $n-$**dimensional vector space**. Then V is **finitely generated** and therefore every **subspace** W of V has a **basis** and is also finite dimensional. Since a basis of W is a **linearly independent** sequence which belong to V we can conclude the following:

Theorem 5.3.7. *Let V be a **vector space** with $dim(V) = n$ and let W be a **subspace** of V. Then the following hold:*

1. $dim(W) \leq n$.

2. $dim(W) = n$ if and only if $W = V$.

Example 5.3.5. Let V be a **vector space** , $dim(V) = 3$. Let $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ be a sequence of vectors from V.

1. Show that if S is **linearly independent** then S **spans** V and so is a **basis** of V.

2. Show that if S **spans** V then S is **linearly independent** and so is a **basis** of V.

1. Suppose to the contrary that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) \neq V$ and that $\boldsymbol{u}$ is a vector in V and $\boldsymbol{u}$ is not in $Span(S)$. Then by part 2) of **Theorem** (5.2.7) the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{u})$ is **linearly independent** which contradicts the **Exchange Theorem**. Thus, $S-(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ **spans** V as claimed.

2. Suppose to the contrary that S is **linearly dependent**. By part 3) of **Theorem** (5.2.5) one of the vectors (say $\boldsymbol{v}_3$) can be expressed as a **linear combination** of the remaining vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2)$. By **Theorem** (5.2.4) $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = V$. However, this contradicts part 2) of **Theorem** (5.3.7).

This same reasoning is used to prove the following general result which is the "Half is good enough theorem" for abstract **vector spaces**:

Theorem 5.3.8. *Let V be* **n-dimensional vector space** *and $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ be a sequence of vectors from V. Then the following hold:*

1. If S is **linearly independent** *then S* **spans** *V and S is a* **basis** *of V.*

2. If S **spans** *V then S is* **linearly independent** *and S is a* **basis** *of V.*

More succinctly, we can say that S **spans** *V if and only if S is* **linearly independent** *if and only if S is a* **basis** *of V.*

Proof. 1. We do a **proof by contradiction**. Suppose S does not **span** V. Then there is a vector $\boldsymbol{u} \in V, \boldsymbol{u} \notin Span(S)$. But then by part 2) of **Theorem** (5.2.7) $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n, \boldsymbol{u})$ is **linearly independent**. However, by the **Exchange Theorem**, it is not possible for a **linearly independent** sequence to have length $n+1$ and we have a contradiction. Therefore S **spans** V and is a **basis** for V.

2. This is left as a **challenge exercise**. $\square$

Example 5.3.6. Let V be the **subspace** of $\mathbb{R}^4$ of those vectors $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ such that

$$x_1 + x_2 + x_3 + x_4 = 0.$$

We claim that the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a **basis** where

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

First of all if a **linear combination**

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ -(c_1 + c_2 + c_3) \end{pmatrix} = \mathbf{0}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

then $c_1 = c_2 = c_3$ and so the sequence is **linearly independent**.

On the other hand, suppose $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ is in V. Then $x_1 + x_2 + x_3 + x_4 = 0$ which implies that $x_4 = -(x_1 + x_2 + x_3)$. As a consequence,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_2 \\ x_4 \end{pmatrix} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + x_3\boldsymbol{v}_3.$$

Therefore, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ also **spans** V and hence $dim(V) = 3$.

Now set

$$\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -3 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 2 \\ 3 \\ -1 \\ -4 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 3 \\ 4 \\ 1 \\ -8 \end{pmatrix}.$$

We will prove that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is a **basis** for V.

First, it is immediate that each vector belongs to V since the sum of its components is 0:

$$1 + 1 + 1 - 3 = 0$$

$$2 + 3 - 1 - 4 = 0$$

$$3 + 4 + 1 - 8 = 0.$$

By the **Half is Good Enough Theorem** we need only show that the sequence is **linearly independent**. We have to make a matrix with the vectors as columns, use **Gaussian elimination** to obtain an **echelon form** and determine if every column is a **pivot column**.

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & -1 & 1 \\ -3 & -4 & -8 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & -3 & -2 \\ 0 & 2 & 1 \end{pmatrix} \to$$

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

The last matrix is an **echelon form** with three pivots and therefore the sequence $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent** and consequently a **basis** of V.

We state and prove one last result:

Theorem 5.3.9. *Let V be an* **n-dimensional vector space** *and $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ a sequence of vectors from V.*

1. If S is **linearly independent** *and $m < n$ then S can be expanded to a* **basis** *for V, that is, there are vectors $\boldsymbol{v}_{m+1}, \dots, \boldsymbol{v}_n$ such that $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ is a basis of V.*

2. If S **spans** *V and $m > n$ then S can be contracted to a* **basis** *of V, that is, there exists a subsequence $S' = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_n)$ which is a basis of V.*

Proof.

1. Since $m < n$ it follows that S is not a **basis** of V. Since S is **linearly independent** we can conclude that S does not **span** V. Therefore there is a vector $\boldsymbol{v}$ which is not in $Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_m)$. Set $\boldsymbol{v}_{m+1} = \boldsymbol{v}$. By **Theorem** (5.2.7) $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_m, \boldsymbol{v}_{m+1})$ is **linearly independent**. Among all **linearly independent** sequences $(\boldsymbol{v}_{m+1}, \dots, \boldsymbol{v}_l)$ choose $l \leq n$ as large as possible such that $T = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_m, \boldsymbol{v}_{m+1}, \dots, \boldsymbol{v}_l)$ is linearly independent. We claim that the sequence T **spans** V and is therefore a **basis** of V. We do a **proof by contradication**.

Assume to the contrary that T does not **span** V. Then there is a vector $\boldsymbol{u}$ in V which is not in $Span(T) = Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_m, \boldsymbol{v}_{m+1}, \dots, \boldsymbol{v}_l)$. Set $\boldsymbol{v}_{l+1} = \boldsymbol{u}$. By **Theorem** (5.2.7) the sequence $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_m, \boldsymbol{v}_{m+1}, \dots, \boldsymbol{v}_l, \boldsymbol{v}_{l+1})$ is **linearly independent**. However, this contradicts the maximality of l.

2. Since $m > n$ it follows that S is not a **basis** of V. Since S **spans** V we can conclude that S is **linearly dependent**. By part 3) of **Theorem** (5.2.5) some vector $\boldsymbol{v}_i$ of S is a **linear combination** of the sequence S^* obtained when $\boldsymbol{v}_i$ is deleted. Then by **Theorem** (5.2.4) $Span(S^*) = Span(S) = V$. This proves that there are proper subsequences of S which **span** V.

Now let S' to be a subsequence of S such that $Span(S') = Span(S) = V$ and chosen so that the length of S' is as small as possible. We claim that S' is **linearly independent** and therefore a **basis** of V. Again, we do a **proof by contradication**.

Assume to the contrary that S' is **linearly dependent**. Then by part 3) of **Theorem** (5.2.5) some vector $\boldsymbol{v}_j$ of S' is a **linear combination** of the sequence $S^\star$ obtained when $\boldsymbol{v}_j$ is deleted. Then by **Theorem** (5.2.4) $Span(S^\star) = Span(S') = V$. However, the length of $S^\star$ is less than the length of S' and we have a contradiction. □

What You Can Now Do

1. Compute the **dimension** of a **finitely generated vector space**.

2. Given a sequence $\mathcal{B}$ of vectors in an **n-dimensional vector space** determine if it is a **basis**.

Methods (How To Do It)

Method 5.3.1. Compute the **dimension** of a **finitely generated vector space**.

In general to do this you need to find a **basis** and compute its length. However, sometimes this can be simplified when you have additional information.

Example 5.3.7. Let V consist of all those vectors $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$ in **5-space** which satisfy

$$x_1 - 2x_2 + 3x_3 - 4x_4 + 5x_5 = 0.$$

Find the **dimension** of V.

First note that V is a **subspace** as can be shown using the methods of Section (2.5). Also, $V \neq \mathbb{R}^5$ since, for example, $\boldsymbol{e}_1^5 \notin V$. Therefore $dim(V) \neq 5$ and consequently, $dim(V) \leq 4$. If we can exhibit a **linearly independent** sequence with four vectors in V then we will also know that $dim(V) \geq 4$ and therefore $dim(V) = 4$.

Consider the following four vectors:

$$\boldsymbol{v}_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} -3 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} -5 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

By inspection $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4 \in V$. We claim the sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is **linearly independent**. For suppose c_1, c_2, c_3, c_4 are scalars and $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 = \boldsymbol{0}_5$. We then have

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 = \boldsymbol{0}_5 = \begin{pmatrix} 2c_1 - 3c_2 + 4c_3 - 5c_4 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus, $c_1 = c_2 = c_3 = c_4 = 0$ as required.

We remark that the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is a **basis** for V by **Theorem** (5.3.8).

Example 5.3.8. Let $A =$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 3 & 2 & 2 & 3 & 4 \\ 5 & 3 & 2 & 0 & 5 & 6 \\ 2 & 3 & 4 & 3 & 3 & 3 \\ 7 & 3 & 0 & -1 & 5 & 7 \end{pmatrix}$$

Find the **dimension** of the **null space** of A.

We compute **null(A)** by **Method** (3.2.2) and as we will see that this gives a **basis** for $null(A)$. Thus, we use **Gaussian elimination** to get the **reduced echelon form** of A:

$$\to \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -2 & -1 & 0 \\ 0 & -2 & -3 & -5 & 0 & 1 \\ 0 & 1 & 2 & 1 & 1 & 1 \\ 0 & -4 & -7 & -8 & -2 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -2 & -1 & 0 \\ 0 & 0 & 1 & -1 & 2 & 1 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 & 0 \end{pmatrix} \to$$

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

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

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

Now we write out the **homogeneous linear system** with this matrix as its **coefficient matrix**:

$$\begin{array}{ccccccc} x_1 & & & & +2x_5 & & = 0 \\ & x_2 & & & -3x_5 & +2x_6 & = 0 \\ & & x_3 & & +2x_5 & & = 0 \\ & & & x_4 & & -x_6 & = 0 \end{array}$$

There are four **leading variables** (x_1, x_2, x_3, x_4) and two **free variables** $(x_5, x_6.)$ We set $x_5 = s, x_6 = t$ and express each variable in terms of s, t :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} -2s \\ 3s-2t \\ -2s \\ t \\ s \\ t \end{pmatrix} = \begin{pmatrix} -2s \\ 3s \\ -2s \\ 0 \\ s \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -2t \\ 0 \\ t \\ 0 \\ t \end{pmatrix} = s\begin{pmatrix} -2 \\ 3 \\ -2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

This shows that the **null space** of A is **spanned** by $\left(\begin{pmatrix} -2 \\ 3 \\ -2 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right)$.

The only way a **linear combination** is the zero vector is when $s = t = 0$. Thus,

$$\left(\begin{pmatrix} -2 \\ 3 \\ -2 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right)$$ is a **basis** and the **dimension** of **null(A)** is 2.

Remark 5.7. In general if A is a matrix then **dimension** of **null(A)** is the number of **non-pivot columns** of an **echelon form** of A. We will return to this in the next section.

Example 5.3.9. Let $\boldsymbol{m}(x, y, z) = \begin{pmatrix} x + 3y + 2z & 3x + y - 2z \\ 2x + 5y + 3z & 2x + 3y + z \end{pmatrix}$.

Set $S = \{\boldsymbol{m}(x, y, z) \in M_{2\times 2}(\mathbb{R}) : x, y, z \in \mathbb{R}\}$. Find **dim(S)**.

By the methods of Section (5.1) we can show that S is a subspace of $M_{22}(\mathbb{R})$.

The typical element $\boldsymbol{m}(x, y, z)$ of S can be expressed as

$$\boldsymbol{m}(x, y, z) = x\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix} + y\begin{pmatrix} 3 & 1 \\ 5 & 3 \end{pmatrix} + z\begin{pmatrix} 2 & -2 \\ 3 & 1 \end{pmatrix}.$$

Thus, we conclude that

$$S = Span\left(\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 5 & 3 \end{pmatrix}, \begin{pmatrix} 2 & -2 \\ 3 & 1 \end{pmatrix}\right)$$

.
Consequently, $dim(S) \leq 3$. If this sequence of is **linearly independent** then it will be a **basis** and **dim(S)** = 3. We use **Method** (5.2.3) to determine if the sequence is **linearly independent**. Thus, associate to each matrix of S a **4-vector**. We then make these the columns of a matrix and use **Gaussian elimination** to find an **echelon form**. The required matrix is

$$\begin{pmatrix} 1 & 3 & 2 \\ 2 & 5 & 3 \\ 3 & 1 & -2 \\ 2 & 3 & 1 \end{pmatrix}$$

We apply **Gaussian elimination**:

$$\to \begin{pmatrix} 1 & 3 & 2 \\ 0 & -1 & -1 \\ 0 & -8 & -8 \\ 0 & -3 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So we see the sequence is **linearly dependent** and the last vector of the sequence is a **linear combination** of the two proceding vectors. Thus, S can be **spanned** by a sequence of length 2. Since $\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix}$ and $\begin{pmatrix} 3 & 1 \\ 5 & 3 \end{pmatrix}$ are not multiples of one another, $\left(\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 5 & 3 \end{pmatrix}\right)$ is **linearly independent** and we can conclude that **dim(S)** is 2.

Method 5.3.2. Given a sequence, $\mathcal{B}$, of vectors in an **n-dimensional vector space** determine if it is a **basis**.

If the length of $\mathcal{B}$ is less than n vectors we can conclude that $\mathcal{B}$ does not **span** V and therefore is not a **basis** of V. If the length of $\mathcal{B}$ is greater than n vectors in $\mathcal{B}$ we can conclude that $\mathcal{B}$ is **linearly dependent** and, consequently, is not a **basis** of V.

If the length of $\mathcal{B}$ is n determine whether $\mathcal{B}$ **spans** V or is **linearly independent**. If either, then it is a **basis** of V, otherwise it is not.

Example 5.3.10. Let

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \boldsymbol{m}_2 \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$\boldsymbol{m}_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \boldsymbol{m}_5 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

Determine if $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5)$ is a **basis** of $M_{2\times 3}(\mathbb{R})$.

The **dimension** of $M_{2\times 3}(\mathbb{R})$ is six. The length of the sequence is 5 and we therefore we conclude that this is not a **basis**.

Example 5.3.11. We continue with the last example. Determine if the sequence obtained by adjoining the matrix

$$\boldsymbol{m}_6 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

is a basis for $M_{2\times 3}(\mathbb{R})$.

It is not difficult to see that the sequence of the first five matrices is **linearly independent** since a typical **linear combination** is $\begin{pmatrix} a & b & c \\ d & e & -(a+b+c+d+e) \end{pmatrix}$ and this is the zero vector, $\mathbf{0}_{2\times 3}$, if and only if $a = b = c = d = e = 0$.

We can also see that for any matrix in $W = Span(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5)$ the sum of its entries is 0. Consequently, $\boldsymbol{m}_6$ is not in W. By part 2) of **Theorem** (5.2.7) the sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5, \boldsymbol{m}_6)$ is **linearly independent**. By **Theorem** (5.3.8) it is a **basis** of $M_{2\times 3}(\mathbb{R})$.

Example 5.3.12. Determine if the sequence $(1, 1+2x+x^2, 1-2x+x^2, 1+4x+4x^2)$ a **basis** of $\mathbb{R}_2[x]$?

The **dimension** of $\mathbb{R}_2[x]$ is 3. The sequence has length 4 and therefore it is **linearly dependent** and not a **basis** of $\mathbb{R}_2[x]$.

Is the sequence $(1+2x+x^2, 1-2x+x^2, 1+4x+4x^2)$ a basis for $\mathbb{R}_2[x]$?

We need only check and see if the sequence **spans** $\mathbb{R}_2[x]$ or is **linearly independent** using either **Method** (5.2.1) or **Method** (5.2.3).

We associate to each vector a **3-vector** and make these the columns of a matrix. The matrix we obtain is

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & -2 & 4 \\ 1 & 1 & 4 \end{pmatrix}$$

We now use **Gaussian elimination** to obtain an **echelon form**:

$$\to \begin{pmatrix} 1 & 1 & 1 \\ 0 & -4 & 2 \\ 0 & 0 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}.$$

Every column is a **pivot column** so we conclude the sequence is **linearly independent**. Since the length of the sequence is three and the **dimension** of $\mathbb{R}_2[x]$ is three, **Theorem** (5.3.8) applies and the sequence is a **basis** of $\mathbb{R}_2[x]$.

Exercises

For exercises 1-6 see **Method** (5.3.1).

1. Let $V = Span(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5)$ where

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{m}_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \boldsymbol{m}_5 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Compute the **dimension** of V.

2. Let $\boldsymbol{f}_1 = 1 + x + x^4$, $\boldsymbol{f}_2 = 1 + x^2 + x^4$, $\boldsymbol{f}_3 = 1 + x^3 + x^4$,

$\boldsymbol{f}_4 = 1 + x + x^2 + x^3 + 2x^4$ and $\boldsymbol{f}_5 = 1 + 2x + 3x^2 + 4x^4 + 5x^4$.

Set $S = Span(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4, \boldsymbol{f}_5)$. Compute the **dimension** of S.

3. Let $\boldsymbol{m}(a,b,c,d) = \begin{pmatrix} a+b+c & a+2b+4c+2d \\ a+2b+3c+d & a+2b+3c+2d \end{pmatrix}$ and set $M = \{\boldsymbol{m}(a,b,c,d) \in M_{2\times 2}(\mathbb{R}) \mid a,b,c,d \in \mathbb{R}\}$. Compute the **dimension** of M.

4. Let $\boldsymbol{f}(a,b,c,d) = (a+b+c+d) + (2a+3b+4d)x + (-a+b-5c+3d)x^2 + (3a+6b-3c+9d)x^3$ and set $P = \{\boldsymbol{f}(a,b,c,d) \in \mathbb{R}_3[X] : a,b,c,d \in \mathbb{R}\}$. Compute the **dimension** of P.

5. Let $A = \begin{pmatrix} 1 & 2 & 0 & -1 & 2 \\ 3 & 7 & 2 & -2 & 9 \\ 2 & 3 & -2 & -3 & 1 \\ 2 & 5 & 2 & -2 & 2 \\ 1 & 3 & 2 & 0 & 5 \end{pmatrix}$. Compute the **dimension** of the **null space** of A, $dim(null(A))$.

6. Let $A = \begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 7 & 5 & 3 \\ 1 & 2 & 2 & 2 & -1 \\ 3 & 6 & 5 & 6 & -4 \end{pmatrix}$ Compute the **dimension** of the **null space** of A, $dim(null(A))$.

7. Let $S = \{a_0 + a_1x + a_2x^2 + a_3x^3 \in \mathbb{R}_3[x] : a_0 + a_1 + a_2 + 3a_3 = 0\}$. Verify that $\mathcal{B} = (-1+x, -1+x^2, -3+x^3)$ is a **basis** of S.

In exercises 8 - 12 decide if the sequence $\mathcal{B}'$ is a **basis** for the space S of exercise 7.

See **Method** (5.3.2). Note that if there are the right number of vectors you still have to show that the vectors belong to the **subspace** S.

8. $\mathcal{B}' = (1 + x + x^2 - x^3, 1 + 2x - x^3, x + 2x^2 - x^3)$.

9. $\mathcal{B}' = (3 - x^3, 3x - x^3, 3x^2 - x^3)$.

10. $\mathcal{B}' = (1 + 2x + 3x^2 - 2x^3, 2 + 3x + x^2 - 2x^3)$.

11. $\mathcal{B}' = (1 + x - 2x^2, -2 + x + x^2, 1 - 2x + x^2, 1 + x + x^2 - x^3)$.

12. $\mathcal{B}' = (1 + x - 2x^2, 3 + 2x + x^2 - 2x^3, 2 + x + 3x^2 - 2x^3)$.

13. Let $T = \{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \in M_{2\times 3}(\mathbb{R}) : a_{11} + a_{12} + a_{13} = a_{21} + a_{22} + a_{23} = 0\}$.

Verify that $\mathcal{B} = \left(\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}\right)$ is a **basis** for T.

In exercises 14-19 determine if the sequence $\mathcal{B}'$ is a **basis** for the space T of exercise 13.

See **Method** (5.3.2). Note that if there are the right number of vectors you still have to show that the vectors belong to the **subspace** S.

14. $\mathcal{B}' = \left(\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix} \right)$.

15. $\mathcal{B}' = \left(\begin{pmatrix} 1 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & -2 \end{pmatrix}, \begin{pmatrix} 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix} \right)$.

16. $\mathcal{B}' = \left(\begin{pmatrix} 1 & -1 & 0 \\ 1 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 2 & -3 \\ 0 & 0 & 0 \end{pmatrix} \right)$.

17. $\mathcal{B}' = \left(\begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \end{pmatrix}, \begin{pmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \end{pmatrix} \right)$.

18. $\mathcal{B}' = \left(\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \end{pmatrix} \right)$.

19. $\mathcal{B}' = \left(\begin{pmatrix} 1 & 1 & -2 \\ 1 & 1 & -2 \end{pmatrix}, \begin{pmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \right)$.

In exercises 20 - 24 answer true or false and give an explanation.

20. If V is a **vector space** of **dimension** three and W is a **subspace** of V then the dimension of W is two.

21. If W is a **subspace** of a **vector space** V, the **dimension** of W is three, and $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent** from V then $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is a **basis** of W.

22. A **basis** for $\mathbb{R}^5$ can be contracted to a basis of $\mathbb{R}^4$.

23. A **basis** of $M_{2\times 2}(\mathbb{R})$ can be extended to a basis of $M_{2\times 3}(\mathbb{R})$.

24. A **basis** of $\mathbb{R}_2[x]$ can be extended to a basis of $\mathbb{R}_3[x]$.

Challenge Exercises (Problems)

1. Assume that U and W are **subspaces** of a **four dimensional vector space** V, $U \neq W$, and $dim(U) = dim(W) = 3$. Prove that $U + W = V$.

Recall that $U + W$ consists of all vectors $\boldsymbol{u} + \boldsymbol{v}$ where $\boldsymbol{u} \in U, \boldsymbol{v} \in W$.

2. Assume that U and W are **subspaces** of a **vector space** V and that $U \cap W = \{\mathbf{0}_V\}$. Assume that $(\boldsymbol{u}_1, \boldsymbol{u}_2)$ is a **basis** for U and $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is a basis for W. Prove that $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is a **basis** for $U + W$.

3. Let $S = \{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} : a_{11}+a_{22} = 0\}$. Prove that there are no **linearly independent** sequences of length 4 in S.

4. Let U and W be **subspaces** of a **finitely generated vector space** V. Prove that $dim(U + W) = dim(U) + dim(W) - dim(U \cap W)$. Hint: Start with a **basis** $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ for $U \cap W$. You can extend by some sequence $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_l)$ of U to get a **basis** $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{u}_1, \dots, \boldsymbol{u}_l)$ of U. You can also extend by some sequence $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_m)$ of W to get a **basis** $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{w}_1, \boldsymbol{w}_m)$ for W. Prove that $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k, \boldsymbol{u}_1, \dots, \boldsymbol{u}_l, \boldsymbol{w}_1, \dots, \boldsymbol{w}_m)$ is a **basis** of $U + W$.

5. Prove the second part of **Theorem** (5.3.8).

6. Prove **Theorem** (5.3.3).

Quiz Solutions

1. The following sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4, \boldsymbol{m}_5)$ is **linearly dependent** because it contains repeated vectors.

Not right, see **Theorem** (5.2.6)

2. The sequence $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, \boldsymbol{f}_4)$ is **linearly dependent**.

Not right, see **Method** (5.2.3).

3. The sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is not a **spanning sequence** of $M_{2\times 2}(\mathbb{R})$.

Not right, see **Method** (5.2.3).

4. The sequence $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ of vectors is a **basis** of $M_{2\times 2}(\mathbb{R})$. T

Not right, see **Method** (5.2.6).

5. The sequence $(1 - x, 1 - x^2, 1 - x^3, 1 + x^3, 1 + x + x^2 + x^3)$ is not a **basis** of $\mathbb{R}_3[x]$ because the sequence is **linearly dependent**. Specifically, a linearly independent sequence of $\mathbb{R}_3[x]$ cannot have length greater than 4 four.

Not right, see **Method** (5.2.3).

5.4. Coordinate vectors and change of basis

The notion of a coordinate vector is defined as well as the change of basis matrix which relates coordinate vectors with respect to different bases of the same vector space.

Am I Ready for This Material

Readiness Quiz

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

The following concepts and procedures are essential for understanding this module:

vector space

span of a sequence of vectors

linearly dependent sequence of vectors

linearly independent sequence of vectors

spanning sequence of a vector space

basis of a vector space

finitely generated vector space

dimension of a finitely generated vector space

finite dimensional vector space

Some procedures that will be used repeatedly in this section are"

Gaussian elimination

matrix inversion algorithm

Quiz

Let $\boldsymbol{f}_1 = 1 + x + 2x^2, \boldsymbol{f}_2 = 3 + 2x + x^2, \boldsymbol{f}_3 = 1 - 2x^2$.

1. Prove that $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ is a **basis** of $\mathbb{R}_2[x]$.

2. Express each of the vectors of the **standard basis** $(1, x, x^2)$ of $\mathbb{R}_2[x]$ as a **linear combination** of $(\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$.

Let

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 2 & 2 \\ 3 & 2 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}, \boldsymbol{m}_4 = \begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix}.$$

3. Prove that $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$.

4. Express the vector

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

as a **linear combination** of $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$.

Quiz Solutions

New Concepts

Two new extremely inmportant concepts are introduced:

coordinate vector of v with respect to a basis

change of basis matrix from one basis to another

Theory (Why It Works)

Recall the following characterization of **bases**:

Theorem (5.2.9) A sequence $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ of a **vector space** V is a **basis** of V if and only if for each vector $\boldsymbol{v}$ in V there is a unique sequence of scalars $(c_1, c_2, \dots, c_k)$ such that

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k.$$

We do an example in $\mathbb{R}^3$:

Example 5.4.1. The most obvious basis in $\mathbb{R}^3$ is the **standard basis**, $\mathcal{S} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. The matrix with these vectors as columns is, of course, the **identity matrix**, $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Since $\mathcal{S}$ is **basis** , as the theorem says, every vector $\boldsymbol{v} \in \mathbb{R}^3$ is uniquely a **linear combination** of $\mathcal{S}$. For this **basis** determining the **linear combination** is very easy: For example,

$$\boldsymbol{v} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = (-1)\boldsymbol{e}_1 + 1\boldsymbol{e}_2 + 0\boldsymbol{e}_3.$$

However, the theorem applies to every **basis** of $\mathbb{R}^3$ not just the $\mathcal{S}$.

Example 5.4.2. The sequence consisting of the three vectors $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, $\boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix}$ is also a **basis** of $\mathbb{R}^3$.

We can also write $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ uniquely as a **linear combination** of $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$:

$$\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \boldsymbol{v}_1 + \boldsymbol{v}_2 - \boldsymbol{v}_3.$$

Such expressions are very important and a useful tool for both theory and computation. For further reference we assign a special name for the vectors which arise this way.

Definition 5.16. Let $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a **basis** for the **vector space** V and let $\boldsymbol{v}$ be a vector in V. If the unique expression of $\boldsymbol{v}$ as a **linear combination** of $\mathcal{B}$ is $\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$ then the vector with $c_1, \ldots, c_n$ as its components, $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$ is called the ***coordinate vector*** of $\boldsymbol{v}$ **with respect to** $\mathcal{B}$. It is denoted by $[\boldsymbol{v}]_{\mathcal{B}}$.

Remark 5.8. In general, for different bases, $\mathcal{B}$, $[\boldsymbol{v}]_{\mathcal{B}}$ will be different. In particular, if the order of $\mathcal{B}$ is changed then $[\boldsymbol{v}]_{\mathcal{B}}$ will change.

Example 5.4.3. Let $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ and $\boldsymbol{v}$ be as introduced in **Example** (5.4.2). Then as we saw, $[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

On the other hand, if $\mathcal{B}' = (\boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{v}_2, \boldsymbol{v}_3)$ then $[\boldsymbol{v}]_{\mathcal{B}'} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$.

If $\mathcal{B}^* = (\boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_1)$ then $[\boldsymbol{v}]_{\mathcal{B}^*} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.

Example 5.4.4. Let $\boldsymbol{f}_1(x) = \frac{1}{2}(x-1)(x-2), \boldsymbol{f}_2(x) = -x(x-2)$ and $\boldsymbol{f}_3(x) = \frac{1}{2}x(x-1)$. Then $\mathcal{B} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$ is a **basis** for $\mathbb{R}_2[x]$, the vector space of all polynomials of degree at most two. This **basis** is special: For an arbitrary polynomial $\boldsymbol{g}(x) \in \mathbb{R}_2[x]$, $[\boldsymbol{g}]_{\mathcal{B}} = \begin{pmatrix} \boldsymbol{g}(0) \\ \boldsymbol{g}(1) \\ \boldsymbol{g}(2) \end{pmatrix}$.

As a concrete example, if $\boldsymbol{g}(x) = x^2 - x + 1$ then $\boldsymbol{g}(0) = 1, \boldsymbol{g}(1) = 1, \boldsymbol{g}(2) = 3$. We compute:

$$\boldsymbol{f}_1(x)+\boldsymbol{f}_2(x)+3\boldsymbol{f}_3(x) = \frac{1}{2}(x-1)(x-2)-x(x-2)+\frac{3}{2}x(x-1) = x^2-x+1 = \boldsymbol{g}(x).$$

Thus, $[\boldsymbol{g}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$ as claimed.

The next theorem indicates why **coordinate vectors** are such a powerful tool: they allow us to translate questions from an abstract vector space to $\mathbb{R}^n$ where we can use the algorithms we have already developed.

Theorem 5.4.1. *Let V be a* **finite dimensional vector space** *with* **basis** *$\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$. Suppose $\boldsymbol{w}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ are vectors in V. Then $\boldsymbol{w}$ is a* **linear combination** *of $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k)$ if and only if $[\boldsymbol{w}]_{\mathcal{B}}$ is a linear combination of $([\boldsymbol{u}_1]_{\mathcal{B}}, [\boldsymbol{u}_2]_{\mathcal{B}}, \ldots, [\boldsymbol{u}_k]_{\mathcal{B}})$.*

More precisely, $\boldsymbol{w} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_k\boldsymbol{u}_k$ and if only if $[\boldsymbol{w}]_{\mathcal{B}} = c_1[\boldsymbol{u}_1]_{\mathcal{B}} + c_2[\boldsymbol{u}_2]_{\mathcal{B}} + \cdots + c_k[\boldsymbol{u}_k]_{\mathcal{B}}$.

Proof. For illustration purposes we do the case where $n = 3, k = 2$.

Suppose $[\boldsymbol{w}]_{\mathcal{B}} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}, [\boldsymbol{u}_1]_{\mathcal{B}} = \begin{pmatrix} u_{11} \\ u_{21} \\ u_{31} \end{pmatrix}, [\boldsymbol{u}_2]_{\mathcal{B}} = \begin{pmatrix} u_{12} \\ u_{22} \\ u_{32} \end{pmatrix}$.

This means: $\boldsymbol{w} = w_1\boldsymbol{v}_1 + w_2\boldsymbol{v}_2 + w_3\boldsymbol{v}_3, \boldsymbol{u}_1 = u_{11}\boldsymbol{v}_1 + u_{21}\boldsymbol{v}_2 + u_{31}\boldsymbol{v}_3$, and $\boldsymbol{u}_2 = u_{12}\boldsymbol{v}_1 + u_{22}\boldsymbol{v}_2 + u_{32}\boldsymbol{v}_3$. Now suppose $\boldsymbol{w} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2$.

Then $\boldsymbol{w} = c_1(u_{11}\boldsymbol{v}_1 + u_{21}\boldsymbol{v}_2 + u_{31}\boldsymbol{v}_3) + c_2(u_{12}\boldsymbol{v}_1 + u_{22}\boldsymbol{v}_2 + u_{32}\boldsymbol{v}_3) = (c_1u_{11} + c_2u_{12})\boldsymbol{v}_1 + (c_1u_{21} + c_2u_{22})\boldsymbol{v}_2 + (c_1u_{31} + c_2u_{32})\boldsymbol{v}_3$.

Thus, $[\boldsymbol{w}]_{\mathcal{B}} = \begin{pmatrix} c_1u_{11} + c_2u_{12} \\ c_1u_{21} + c_2u_{22} \\ c_1u_{31} + c_2u_{32} \end{pmatrix}$.

So, $[\boldsymbol{w}]_{\mathcal{B}} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} c_1u_{11} + c_2u_{12} \\ c_1u_{21} + c_2u_{22} \\ c_1u_{31} + c_2u_{32} \end{pmatrix} = c_1 \begin{pmatrix} u_{11} \\ u_{21} \\ u_{31} \end{pmatrix} + c_2 \begin{pmatrix} u_{12} \\ u_{22} \\ u_{32} \end{pmatrix} = c_1[\boldsymbol{u}_1]_{\mathcal{B}} + c_2[\boldsymbol{u}_2]_{\mathcal{B}}$.

It is straightforward to reverse the argument. The general argument follows exactly this line of reasoning. □

Applying the theorem when $\boldsymbol{w} = \mathbf{0}_V$ we get the following result:

Theorem 5.4.2. *Let V be a* **finite dimensional vector space** *with* **basis** *$\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$. Let $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_k)$ be a sequence of vectors in V. Then $(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k)$ is* **linearly dependent** *if and only if $([\boldsymbol{u}_1]_{\mathcal{B}}, [\boldsymbol{u}_2]_{\mathcal{B}}, \dots, [\boldsymbol{u}_k]_{\mathcal{B}})$ is* **linearly dependent** *in $\mathbb{R}^n$. In fact, $(c_1, \dots, c_k)$ is a* **dependence relation** *on $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_k)$ if and only if $(c_1, \dots, c_k)$ is a* **dependence relation** *on $([\boldsymbol{u}_1]_{\mathcal{B}}, \dots, [\boldsymbol{u}_k]_{\mathcal{B}})$.*

Example 5.4.5. Let $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ and $\boldsymbol{v}$ be as introduced **Example** (5.4.2).

Also set $\boldsymbol{v}_1' = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2' = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}, \boldsymbol{v}_3' = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$.

It is easy to verify that $\mathcal{B}' = (\boldsymbol{v}_1', \boldsymbol{v}_2', \boldsymbol{v}_3')$ is also a **basis** of $\mathbb{R}^3$.

Then we can write $\boldsymbol{v}$ uniquely as a **linear combination** of $\mathcal{B}'$:

$\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = -3\boldsymbol{v}_1' + \boldsymbol{v}_2' + 0\boldsymbol{v}_3'$. We therefore conclude that $[\boldsymbol{v}]_{\mathcal{B}'} = \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}$.

We want to relate $[\boldsymbol{v}]_{\mathcal{B}}$ and $[\boldsymbol{v}]_{\mathcal{B}'}$.

Because $\mathcal{B}'$ is a basis we can write each of the vectors of $\mathcal{B}$ uniquely as a **linear combination** and, of course, these give the **coordinate vectors** of each of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ with respect to $\mathcal{B}'$.

$$[\boldsymbol{v}_1]_{\mathcal{B}'} = \begin{pmatrix} -13 \\ 4 \\ 3 \end{pmatrix}, [\boldsymbol{v}_2]_{\mathcal{B}'} = \begin{pmatrix} -19 \\ 6 \\ 4 \end{pmatrix}, [\boldsymbol{v}_3]_{\mathcal{B}'} = \begin{pmatrix} -29 \\ 9 \\ 7 \end{pmatrix}.$$

Let P be the matrix with these vectors as its columns. We have previously seen that $[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$. Therefore by **Theorem** (5.4.2) it follows that

$$[\boldsymbol{v}]_{\mathcal{B}'} = (1)[\boldsymbol{v}_1]_{\mathcal{B}'} + (1)[\boldsymbol{v}_2]_{\mathcal{B}'} + (-1)[\boldsymbol{v}_3]_{\mathcal{B}'}.$$

Thus,

$$[\boldsymbol{v}]_{\mathcal{B}'} = (1)\times\begin{pmatrix}-13\\4\\3\end{pmatrix} + (1)\times\begin{pmatrix}-19\\6\\4\end{pmatrix} + (-1)\times\begin{pmatrix}-29\\9\\7\end{pmatrix} = P\begin{pmatrix}1\\1\\-1\end{pmatrix} = P[\boldsymbol{v}]_{\mathcal{B}}.$$

Example (5.4.5) illustrates the general case of the following result which we refer to as the ***Change of Basis Theorem***:

Theorem 5.4.3. *Let $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n)$ be two* **bases** *of a* **finite dimensional vector space** *V. Let $P_{\mathcal{B}\to\mathcal{B}'}$ be the matrix whose columns are the coordinate vectors of the $\boldsymbol{v}_i$ with respect to $\mathcal{B}'$: $P_{\mathcal{B}\to\mathcal{B}'} = \left([\boldsymbol{v}_1]_{\mathcal{B}'}[\boldsymbol{v}_2]_{\mathcal{B}'}, \ldots, [\boldsymbol{v}_n]_{\mathcal{B}'}\right)$. Then $P_{\mathcal{B}\to\mathcal{B}'}$ is* **invertible**. *Moreover, for any vector $\boldsymbol{v} \in V$, $[\boldsymbol{v}]_{\mathcal{B}'} = P_{\mathcal{B}\to\mathcal{B}'}[\boldsymbol{v}]_{\mathcal{B}}$.*

Definition 5.17. The matrix $P_{\mathcal{B}\to\mathcal{B}'}$ of **Theorem** (5.4.3) is called the ***change of basis matrix or transition marix from $\mathcal{B}$ to $\mathcal{B}'$***

Remark 5.9. The change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$ is $P^{-1}_{\mathcal{B}\to\mathcal{B}'}$.

The next theorem says that if we begin with a **basis** $\mathcal{B}$ for the **n-dimensional vector space** V then every **invertible** $n\times n$ matrix is the **change of basis matrix** for a suitable basis $\mathcal{B}'$:

Theorem 5.4.4. *Let $\mathcal{B}$ be a* **basis** *for the* **n-dimensional vector space** *V and let P be an $n \times n$ matrix with columns $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n$. Let $\boldsymbol{u}_j$ be the vector in V with* **coordinate vector** *$\boldsymbol{c}_j$ with respect to $\mathcal{B}$, $[\boldsymbol{u}_j]_{\mathcal{B}} = \boldsymbol{c}_j$. Then $\mathcal{B}' = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n)$ is a* **basis** *for V if and only if P is* **invertible**.

Proof. By **Theorem** (5.4.3) the sequence $(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n)$ is **linearly independent** if and only if the sequence of coordinate vectors, $(\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n)$ is **linearly independent** in $\mathbb{R}^n$, if and only if the matrix P is **invertible**. However, since **dim(V)** = n, by **Theorem** (5.3.8), $\mathcal{B}'$ is a **basis** for V if and only if it is **linearly independent**. $\square$

Example 5.4.6. Let $\mathcal{S} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ be the **standard basis** of $\mathbb{R}^3$.

Set $\mathcal{B} = \left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right)$, which is also a basis of $\mathbb{R}^3$. The **change of basis matrix** from $\mathcal{B}$ to $\mathcal{S}$ is the matrix with the vectors of $\mathcal{B}$ as its columns:

$$P_{\mathcal{B}\to\mathcal{S}} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix}. \tag{5.26}$$

The **change of basis matrix** from $\mathcal{B}$ to $\mathcal{S}$ expresses each of the **standard basis vectors** as a **linear combination** of the vectors in $\mathcal{B}$. This is just the **inverse** of the matrix in (5.26), $P_{\mathcal{S}\to\mathcal{B}} = P_{\mathcal{B}\to\mathcal{S}}^{-1} = \begin{pmatrix} -1 & -2 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$.

For example, $\boldsymbol{e}_1 = (-1)\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + 1\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

Now let $\mathcal{B}' = \left(\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}\right)$.

Then $\mathcal{B}'$ is also a **basis** of $\mathbb{R}^3$. The **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{S}$ is

$$P_{\mathcal{B}'\to\mathcal{S}} = \begin{pmatrix} 1 & 2 & -1 \\ -2 & -3 & 2 \\ 0 & 1 & 1 \end{pmatrix}.$$

The **change of basis matrix** from $\mathcal{S}$ to $\mathcal{B}'$ is the **inverse** of $P_{\mathcal{B}'\to\mathcal{S}}$: $P_{\mathcal{S}\to\mathcal{B}'} =$

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

For example, $[\boldsymbol{e}_2]_{\mathcal{B}'} = \begin{pmatrix} -3 \\ 1 \\ -1 \end{pmatrix}$ which we verify:

$$(-3)\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} + (-1)\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \boldsymbol{e}_2.$$

The properties demonstrated in **Example** (5.4.6) hold in general and is the subject of the next result

Theorem 5.4.5. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two* **bases** *of the* **finite dimensional vector space** *V. Then the* **change of basis matrices** *$P_{\mathcal{B}\to\mathcal{B}'}$ and $P_{\mathcal{B}'\to\mathcal{B}}$ are* **inverses** *of each other.*

Proof. Let $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n), \mathcal{B}' = (\boldsymbol{v}_1', \boldsymbol{v}_2', \dots \boldsymbol{v}_n')$. Let $\boldsymbol{c}_j$ be the j^{th} column of the product $P_{\mathcal{B}'\to\mathcal{B}}P_{\mathcal{B}\to\mathcal{B}'}$. Then

$$\boldsymbol{c}_j = P_{\mathcal{B}'\to\mathcal{B}}P_{\mathcal{B}\to\mathcal{B}'}\boldsymbol{e}_j =$$

$$P_{\mathcal{B}'\to\mathcal{B}}P_{\mathcal{B}\to\mathcal{B}'}[\boldsymbol{v}_j]_{[\mathcal{B}]} =$$

$$P_{\mathcal{B}'\to\mathcal{B}}[\boldsymbol{v}_j]_{\mathcal{B}'} = [\boldsymbol{v}_j]_{[\mathcal{B}]} = \boldsymbol{e}_j.$$

Thus, $P_{\mathcal{B}'\to\mathcal{B}}P_{\mathcal{B}\to\mathcal{B}'} = I_n$, the $n\times n$ **identity matrix**. In a similar fashion $P_{\mathcal{B}\to\mathcal{B}'}P_{\mathcal{B}'\to\mathcal{B}} = I_n$. □

In much the same way we can prove the following:

Theorem 5.4.6. *Let $\mathcal{B}$, $\mathcal{B}'$, and $\mathcal{B}^*$ be three bases of the* **finite dimensional vector space** *V. Then the* **change of basis matrix** *from $\mathcal{B}$ to $\mathcal{B}*$ is the product of the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}^*$ by the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$ (on the right):*

$$P_{\mathcal{B}\to\mathcal{B}^*} = P_{\mathcal{B}'\to\mathcal{B}^*}P_{\mathcal{B}\to\mathcal{B}'}.$$

Proof. We know that $P_{\mathcal{B}\to\mathcal{B}^*}$ is the unique matrix such that $P_{\mathcal{B}\to\mathcal{B}^*}[\boldsymbol{v}]_{\mathcal{B}} = [\boldsymbol{v}]_{\mathcal{B}^*}$. However, $P_{\mathcal{B}'\to\mathcal{B}^*}P_{\mathcal{B}\to\mathcal{B}'}[\boldsymbol{v}]_{\mathcal{B}} = P_{\mathcal{B}'\to\mathcal{B}^*}[\boldsymbol{v}]_{\mathcal{B}'} = [\boldsymbol{v}]_{\mathcal{B}^*}$.

Thus, $P_{\mathcal{B}\to\mathcal{B}^*} = P_{\mathcal{B}'\to\mathcal{B}^*}P_{\mathcal{B}\to\mathcal{B}'}$. □

Example 5.4.7. We return to Example (5.4.6) with $\mathcal{B} = \left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right)$ and $\mathcal{B}' = \left(\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \right)$.

We use **Theorem** (5.4.5) and **Theorem** (5.4.6) to compute the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$.

The **change of basis matrix** from $\mathcal{B}$ to $\mathcal{S}$ is the matrix

$$P_{\mathcal{B}\to\mathcal{S}} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix} \tag{5.27}$$

The **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{S}$ is the matrix

$$P_{\mathcal{B}'\to\mathcal{S}} = \begin{pmatrix} 1 & 2 & -1 \\ -2 & -3 & 2 \\ 0 & 1 & 1 \end{pmatrix} \tag{5.28}$$

By **Theorem** (5.4.5) the **change of basis matrix** from $\mathcal{S}$ to $\mathcal{B}'$ is the inverse of the matrix in (5.28) and therefore

$$P_{\mathcal{S}\to\mathcal{B}'} = P^{-1}_{\mathcal{B}'\to\mathcal{S}} = \begin{pmatrix} -5 & -3 & 1 \\ 2 & 1 & 0 \\ -2 & -1 & 1 \end{pmatrix} \tag{5.29}$$

By **Theorem** (5.4.6) the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$ is the product of the change of basis matrix from $\mathcal{B}$ to $\mathcal{S}$ by the change of basis matrix from $\mathcal{S}$ to $\mathcal{B}'$:

$$\begin{gathered} P_{\mathcal{B}\to\mathcal{B}'} = P_{\mathcal{S}\to\mathcal{B}'}P_{\mathcal{B}\to\mathcal{S}} = \\ P^{-1}_{\mathcal{B}'\to\mathcal{S}}P_{\mathcal{B}\to\mathcal{S}} = \\ \begin{pmatrix} -5 & -3 & 1 \\ 2 & 1 & 0 \\ -2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix} = \\ \begin{pmatrix} -2 & -6 & -1 \\ 1 & 2 & -1 \\ -1 & -3 & 0 \end{pmatrix} \end{gathered}$$

For example, according to this calculation we should have $[\boldsymbol{v}_1]_{\mathcal{B}'} = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}$. We check

$$(-2)\boldsymbol{v}_1' + \boldsymbol{v}_2' + (-1)\boldsymbol{v}_3' =$$

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

$$\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \boldsymbol{v}_1.$$

What You Can Now Do

1. Given a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ for a **vector space** V and a vector $\boldsymbol{v}$ in V, compute the **coordinate vector** of $\boldsymbol{v}$ with respect to $\mathcal{B}$.

2. Given a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ for a **vector space** V and the **coordinate vector** of a vector $\boldsymbol{v} \in V$ with respect to $\mathcal{B}$, determine $\boldsymbol{v}$.

3. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}_1', \boldsymbol{v}_2', \ldots, \boldsymbol{v}_n')$. Compute the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$.

4. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}_1', \boldsymbol{v}_2', \ldots, \boldsymbol{v}_n')$. Given the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$ compute the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$.

5. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$, $\mathcal{B}' = (\boldsymbol{v}_1', \boldsymbol{v}_2', \ldots, \boldsymbol{v}_n')$ and $\mathcal{B}^* = (\boldsymbol{v}_1^*, \boldsymbol{v}_2^*, \ldots, \boldsymbol{v}_n^*)$. Given the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$ and the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}^*$ compute the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}^*$.

6. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}_1', \boldsymbol{v}_2', \ldots, \boldsymbol{v}_n')$. Given the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$ and the **coordinate vector** of $\boldsymbol{v}$ with respect to $\mathcal{B}$ compute the coordinate vector of $\boldsymbol{v}$ with respect to $\mathcal{B}'$.

Method (How To Do It)

Method 5.4.1. Given a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ for a **vector space** V and a vector $\boldsymbol{v}$ in V, compute the **coordinate vector** of $\boldsymbol{v}$ with respect to $\mathcal{B}$.

In such exercises the vector space V will be one of $\mathbb{R}^n, \mathbb{R}_{n-1}[x]$ or $M_{k\times l}(\mathbb{R})$ with $kl = n$. In each case there is a corresponding standard basis. For each vector $\boldsymbol{v}_i$ of the basis $\mathcal{B}$ write out the **coordinate vector** of $\boldsymbol{v}_i$ with respect to the standard basis and make the matrix A with these vectors as its columns. Augment the matrix with the coordinate vector $\boldsymbol{c}$ of $\boldsymbol{v}$ with respect to the standard basis to get the augmented matrix $A' = [A|\boldsymbol{c}]$. Now use **Method** (1.2.4) to find the unique solution to the **linear system** with **augmented matrix** equal to A'. The **solution vector** is the coordinate vector of $\boldsymbol{v}$ with respect to $\mathcal{B}$.

Example 5.4.8. Let $\mathcal{S} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ be the standard basis for $\mathbb{R}^3$. Since for any vector $\boldsymbol{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\boldsymbol{e}_1 + y\boldsymbol{e}_2 + z\boldsymbol{e}_3$ we have $[\boldsymbol{v}]_{\mathcal{S}} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \boldsymbol{v}$.

Example 5.4.9. Let $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right)$. This is a **basis** for $\mathbb{R}^3$. Determine the **coordinate vector** of $\boldsymbol{v} = \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix}$ with respect to $\mathcal{B}$.

We form the matrix $\begin{pmatrix} 1 & 2 & 1 & | & 4 \\ 3 & 5 & 2 & | & 2 \\ 1 & 2 & 2 & | & 1 \end{pmatrix}$. After applying **Gaussian elimination** we obtain the **reduced echelon form** of this matrix:
$\begin{pmatrix} 1 & 0 & 0 & | & -19 \\ 0 & 1 & 0 & | & 13 \\ 0 & 0 & 1 & | & -3 \end{pmatrix}$. Therefore the **coordinate vector** is $[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} -19 \\ 13 \\ -3 \end{pmatrix}$.

Example 5.4.10. Let $\boldsymbol{f}_1 = 1 + 2x + x^2$, $\boldsymbol{f}_2 = 2 + 5x + x^2$, $\boldsymbol{f}_3 = 1 + x + 3x^2$, and set $\mathcal{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3)$, a **basis** of $\mathbb{R}_2[x]$. Find $[1 - 3x + 9x^2]_{\mathcal{F}}$.

Form the matrix $\begin{pmatrix} 1 & 2 & 1 & | & 1 \\ 2 & 5 & 1 & | & -3 \\ 1 & 1 & 3 & | & 9 \end{pmatrix}$

By an application of **Gaussian elimination** we obtain the **reduced echelon form**:

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

Therefore $[1 - 3x + 9x^2]_{\mathcal{F}} = \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}$.

Example 5.4.11. For $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3$ and $\mathcal{F}$ as in Example (5.4.10) compute $[\boldsymbol{f}_1]_{\mathcal{F}}, [\boldsymbol{f}_2]_{\mathcal{F}}$, and $[\boldsymbol{f}_3]_{\mathcal{F}}$.

Since $\boldsymbol{f}_1 = 1\boldsymbol{f}_1 + 0\boldsymbol{f}_2 + 0\boldsymbol{f}_3, \boldsymbol{f}_2 = 0\boldsymbol{f}_1 + 1\boldsymbol{f}_2 + 0\boldsymbol{f}_3, \boldsymbol{f}_3 = 0\boldsymbol{f}_1 + 0\boldsymbol{f}_2 + 1\boldsymbol{f}_3$ it follows that

$$[\boldsymbol{f}_1]_{\mathcal{F}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, [\boldsymbol{f}_2]_{\mathcal{F}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, [\boldsymbol{f}_3]_{\mathcal{F}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Example 5.4.12. Let $\mathcal{M} = \left(\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right)$. Find the coordinate vector of $\boldsymbol{m} = \begin{pmatrix} 9 & 4 \\ 1 & 0 \end{pmatrix}$ with respect to $\mathcal{M}$.

We form the matrix $\begin{pmatrix} 1 & 1 & 1 & 1 & | & 9 \\ 1 & 1 & 2 & 1 & | & 1 \\ 1 & 2 & 1 & 1 & | & 4 \\ 1 & 1 & 1 & 2 & | & 0 \end{pmatrix}$

Using Gaussian elimination we obtained the reduced echelon form: $\begin{pmatrix} 1 & 0 & 0 & 0 & | & 31 \\ 0 & 1 & 0 & 0 & | & -5 \\ 0 & 0 & 1 & 0 & | & -8 \\ 0 & 0 & 0 & 1 & | & -9 \end{pmatrix}$.

Therefore the coordinate vector of $\boldsymbol{m}$ is $\begin{pmatrix} 31 \\ -5 \\ -8 \\ -9 \end{pmatrix}$.

Method 5.4.2. Given a basis $\mathcal{B} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ for a vector space V and the coordinate vector, $[\boldsymbol{v}]_{\mathcal{B}}$, of a vector $\boldsymbol{v} \in V$ with respect to $\mathcal{B}$, determine $\boldsymbol{v}$.

Multiply the components of the given coordinate vector by the corresponding basis elements and add up.

Example 5.4.13. Let $\mathcal{F} = (1 + x + x^2, 1 + 2x, 1 + x^2)$. If $[\boldsymbol{g}]_{\mathcal{F}} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}$ determine the vector $\boldsymbol{g}$.

$\boldsymbol{g} = 2(1 + x + x^2) - 3(1 + 2x) + 4(1 + x^2) = 3 - 4x + 6x^2$.

Example 5.4.14. Let $\mathcal{M} = \left(\begin{pmatrix}1 & 1\\1 & 2\end{pmatrix}, \begin{pmatrix}1 & 2\\1 & 1\end{pmatrix}, \begin{pmatrix}1 & 1\\2 & 1\end{pmatrix}, \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}\right)$. If $[\boldsymbol{m}]_{\mathcal{M}} = \begin{pmatrix}2\\-4\\3\\-5\end{pmatrix}$ determine $\boldsymbol{m}$.

$$\boldsymbol{m} = 2\begin{pmatrix}1 & 1\\1 & 2\end{pmatrix} - 4\begin{pmatrix}1 & 2\\1 & 1\end{pmatrix} + 3\begin{pmatrix}1 & 1\\2 & 1\end{pmatrix} - 5\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} = \begin{pmatrix}-4 & -8\\-1 & -2\end{pmatrix}.$$

Method 5.4.3. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n)$. Compute the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$.

Find the **coordinate vectors** of each $\boldsymbol{v}_j$ with respect to the basis $\mathcal{B}'$ by **Method** (5.4.1). The matrix $([\boldsymbol{v}_1]_{\mathcal{B}'}[\boldsymbol{v}_2]_{\mathcal{B}'} \ldots [\boldsymbol{v}_n]_{\mathcal{B}'})$ is the **change of basis matrix** $P_{\mathcal{B}\to\mathcal{B}'}$ from $\mathcal{B}$ to $\mathcal{B}'$. You don't have to do these as individual problems but as a single problem by adjoining all the columns for the $\boldsymbol{v}_j$ vectors to the matrix with columns the $\boldsymbol{v}'_i$ vectors. When we apply **Gaussian elimination** to this matrix to find the **reduced echelon form** the left hand side will become the **identity matrix** and the right hand side will be the **change of basis matrix**.

Example 5.4.15. Let $\mathcal{S}$ be the **standard basis** for $\mathbb{R}^3$. Find $P_{\mathcal{B}\to\mathcal{S}}$ and $P_{\mathcal{S}\to\mathcal{B}}$ where $\mathcal{B}$ is as in **Example** (5.4.9).

For the **standard basis**, $\mathcal{S}$ of $\mathbb{R}^n$, we have $[\boldsymbol{v}]_{\mathcal{S}} = \boldsymbol{v}$ for any vector $\boldsymbol{v} \in \mathbb{R}^n$. Therefore

$$P_{\mathcal{B}\to\mathcal{S}} = \begin{pmatrix}1 & 2 & 1\\3 & 5 & 2\\1 & 2 & 2\end{pmatrix}.$$

To compute $P_{\mathcal{S}\to\mathcal{B}}$ the relevant matrix is

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

The **reduced echelon form** of this matrix, obtained by **Gaussian elimination** , is

$$\left(\begin{array}{ccc|ccc}1 & 0 & 0 & -6 & 2 & 1\\0 & 1 & 0 & 4 & -1 & -1\\0 & 0 & 1 & -1 & 0 & 1\end{array}\right)$$

Therefore the **change of basis matrix** is

$$P_{\mathcal{S}\to\mathcal{B}} = \begin{pmatrix} -6 & 2 & 1 \\ 4 & -1 & -1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Example 5.4.16. $\mathcal{G} = (1+x, x+x^2, 2+x-2x^2)$ is a **basis** for $\mathbb{R}_2[x]$. Find the **change of basis matrix** $P_{\mathcal{F}\to\mathcal{G}}$ where $\mathcal{F}$ is the **basis** defined in **Example** (5.4.10)

The relevant matrix is $\begin{pmatrix} 1 & 0 & 2 & | & 1 & 3 & 1 \\ 1 & 1 & 1 & | & 2 & 5 & 1 \\ 0 & 1 & -2 & | & 1 & 1 & 3 \end{pmatrix}$.

The **reduced echelon form** of this matrix, obtained by **Gaussian elimination**, is

$$\begin{pmatrix} 1 & 0 & 0 & | & 1 & 1 & 7 \\ 0 & 1 & 0 & | & 1 & 3 & -3 \\ 0 & 0 & 1 & | & 0 & 1 & -3 \end{pmatrix}.$$

Therefore the **change of basis matrix** matrix is $P_{\mathcal{F}\to\mathcal{G}} = \begin{pmatrix} 1 & 1 & 7 \\ 1 & 3 & -3 \\ 0 & 1 & -3 \end{pmatrix}$.

Method 5.4.4. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n)$. Given the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$ compute the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$.

The **change of bases matrices** $P_{\mathcal{B}\to\mathcal{B}'}$ and $P_{\mathcal{B}'\to\mathcal{B}}$ are **inverses** of each other and therefore if we are given one of these we can find the other using the **matrix inversion algorithm**.

To see why this works go to **Theorem** (5.4.5).

Example 5.4.17. Let $\mathcal{M}$ be the **basis** for $M_{2\times 2}(\mathbb{R})$ defined in **Example** (5.4.12). Set $\mathcal{M}' = \left(\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 1 & 3 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$, which is also a **basis** for $M_{2\times 2}\mathbb{R})$.

The **change of basis matrix** from $\mathcal{M}$ to $\mathcal{M}'$ is $P = P_{\mathcal{M}\to\mathcal{M}'} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & -1 \\ 0 & 1 & -2 & 0 \\ 0 & -2 & 1 & 1 \end{pmatrix}$.

The **inverse** of P obtained using the **matrix inversion algorithm** is $P_{\mathcal{M}'\to\mathcal{M}} =$

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

Method 5.4.5. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$, $\mathcal{B}' = (\boldsymbol{v}'_1, \boldsymbol{v}'_2, \dots, \boldsymbol{v}'_n)$ and $\mathcal{B}^* = (\boldsymbol{v}^*_1, \boldsymbol{v}^*_2, \dots, \boldsymbol{v}^*_n)$. Given the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$ and the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}^*$ compute the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}^*$.

Multiply $P_{\mathcal{B}' \to \mathcal{B}^*}$ by $P_{\mathcal{B} \to \mathcal{B}'}$ on the right to obtain $P_{\mathcal{B} \to \mathcal{B}^*}$: $P_{\mathcal{B} \to \mathcal{B}^*} = P_{\mathcal{B}' \to \mathcal{B}^*} P_{\mathcal{B} \to \mathcal{B}'}$.

To see why this works go to **Theorem** (5.4.6).

Example 5.4.18. Let $\mathcal{S}_3$ be the **standard basis** of $\mathbb{R}^3$ and $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right)$ the basis introduced in **Example** (5.4.15). There we showed that $P_{\mathcal{S} \to \mathcal{B}}$, is equal to

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

Now set $\mathcal{B}' = \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$. Then $\mathcal{B}'$ is also a **basis** of $\mathbb{R}^3$. The **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{S}$, $P_{\mathcal{B}' \to \mathcal{S}}$, is equal to

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

Then $P_{\mathcal{B}' \to \mathcal{B}} = P_{\mathcal{S} \to \mathcal{B}} P_{\mathcal{B}' \to \mathcal{S}} =$

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

Method 5.4.6. Let V be a **vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n)$. Given the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$ and the **coordinate vector of $\boldsymbol{v}$ with respect to $\mathcal{B}$** compute the coordinate vector of $\boldsymbol{v}$ with respect to $\mathcal{B}'$.

Multiply the **coordinate vector** $[\boldsymbol{v}]_{\mathcal{B}}$ by the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$, $P_{\mathcal{B}\to\mathcal{B}'}$. The result is $[\boldsymbol{v}]_{\mathcal{B}'}$.

Example 5.4.19. Let $\mathcal{B}$ be the **basis** of $\mathbb{R}^3$ introduced in **Example** (5.4.9). We have previously computed the **change of basis matrix** from $\mathcal{S}$ to $\mathcal{B}$, $P_{\mathcal{S}\to\mathcal{B}}$:

$$P_{\mathcal{S}\to\mathcal{B}} = \begin{pmatrix} -6 & 2 & 1 \\ 4 & -1 & -1 \\ -1 & 0 & 1 \end{pmatrix}.$$

We compute the **coordinate vector** of $\boldsymbol{v} = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}$ with respect to $\mathcal{B}$. We know that $[\boldsymbol{v}]_{\mathcal{S}} = \boldsymbol{v}$. Then

$$[\boldsymbol{v}]_{\mathcal{B}} = P_{\mathcal{S}\to\mathcal{B}}[\boldsymbol{v}]_{\mathcal{S}} = P_{\mathcal{S}\to\mathcal{B}}\boldsymbol{v} =$$

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

$$\begin{pmatrix} -21 \\ 11 \\ 2 \end{pmatrix}.$$

Example 5.4.20. Let $\mathcal{F} = (1 + 2x + x^2, 2 + 5x + x^2, 1 + x + 3x^2)$ be the **basis** of $\mathbb{R}_2[x]$ introduced in **Example** (5.4.10). We saw in that example that $[1-3x+9x^2]_{\mathcal{F}} = \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}$.

If we set $\mathcal{G} = (1 + x, x + x^2, 2 + x - 2x^2)$ then $\mathcal{G}$ is also a **basis** of $\mathbb{R}_2[x]$. The **change of basis matrix** from $\mathcal{F}$ to $\mathcal{G}$ is

$$P_{\mathcal{F}\to\mathcal{G}} = \begin{pmatrix} 1 & -2 & 7 \\ 1 & 5 & -3 \\ 0 & 2 & -3 \end{pmatrix}.$$

It then follows that $[1 - 3x + 9x^2]_{\mathcal{G}} = \begin{pmatrix} 1 & -2 & 7 \\ 1 & 5 & -3 \\ 0 & 2 & -3 \end{pmatrix} \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 27 \\ -17 \\ -13 \end{pmatrix}$.

As a check:

$$27(1+x) - 17(x+x^2) - 13(2+x-2x^2) = 1 - 3x + 9x^2.$$

Exercises

Let $\mathcal{F}_S = (1, x, x^2)$,

$\mathcal{F} = (1+x, 1+x^2, 1+2x-2x^2)$,

$\mathcal{F}' = (1-x, 1-x^2, 1-2x+2x^2)$.

Let $\mathcal{M}_S = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$,

$\mathcal{M} = \left(\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$

$\mathcal{M}' = \left(\begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & -5 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 2 & -5 \end{pmatrix} \right)$.

For exercises 1 - 4 see **Method** (5.4.1).

1. Compute $[3+2x+x^2]_{\mathcal{F}}$

2. Compute $[3+2x+x^2]_{\mathcal{F}'}$

3. Compute $[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}]_{\mathcal{M}}$

4. Compute $[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}]_{\mathcal{M}'}$

For exercises 5 - 13 see **Method** (5.4.3).

5. Compute $P_{\mathcal{F} \to \mathcal{F}_S}$

6. Compute $P_{\mathcal{F}_S \to \mathcal{F}}$

7. Compute $P_{\mathcal{F}_S \to \mathcal{F}'}$

8. Compute $P_{\mathcal{F} \to \mathcal{F}'}$

9. Compute $P_{\mathcal{F}' \to \mathcal{F}}$

10. Compute $P_{\mathcal{M}_S \to \mathcal{M}}$

11. Compute $P_{\mathcal{M}_S \to \mathcal{M}'}$

12. Compute $P_{\mathcal{M} \to \mathcal{M}'}$

13. Compute $P_{\mathcal{M}' \to \mathcal{M}}$

V is a **vector space** with **basis** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Set $\boldsymbol{u}_1 = 2\boldsymbol{v}_1 - \boldsymbol{v}_2 + \boldsymbol{v}_3, \boldsymbol{u}_2 = -\boldsymbol{v}_1 + \boldsymbol{v}_2 + 2\boldsymbol{v}_3, \boldsymbol{u}_3 = -\boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 6\boldsymbol{v}_3$,

14. Let $\mathcal{B}' = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$. Demonstrate that $\mathcal{B}'$ is a **basis** of V. See **Theorem** (5.4.4).

15. If $[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 5 \\ -4 \end{pmatrix}$ compute $[\boldsymbol{v}]_{\mathcal{B}'}$. See **Method** (5.4.6)

Let $[\boldsymbol{m}_1]_{\mathcal{M}} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ -2 \end{pmatrix}, [\boldsymbol{m}_2]_{\mathcal{M}} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ -2 \end{pmatrix}, [\boldsymbol{m}_3]_{\mathcal{M}} = \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \end{pmatrix}, [\boldsymbol{m}_4]_{\mathcal{M}} = \begin{pmatrix} 1 \\ 1 \\ -2 \\ 1 \end{pmatrix}$.

16. Prove that $\mathcal{M}^* = (\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$. See **Theorem** (5.4.4).

17. Compute $P_{\mathcal{M}^* \to \mathcal{M}}$.
See **Method** (5.4.3).

18. Compute $P_{\mathcal{M} \to \mathcal{M}^*}$. See **Method** (5.4.4).

19. If $[\boldsymbol{m}]_{\mathcal{M}} = \begin{pmatrix} 4 \\ 2 \\ 3 \\ 1 \end{pmatrix}$ determine $\boldsymbol{m}$. See **Method** (5.4.2).

20. If $[\boldsymbol{m}]_{\mathcal{M}} = \begin{pmatrix} 4 \\ 2 \\ 3 \\ 1 \end{pmatrix}$ determine $[\boldsymbol{m}]_{\mathcal{M}^*}$. See **Method** (5.4.6).

In exercises 21 - 24 answer true or false and give an explanation. In all these exercises V is a **3-dimensional vector space** with **bases** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ and $\mathcal{B}' = (\boldsymbol{v}'_1, \boldsymbol{v}'_2, \boldsymbol{v}'_3)$.

21. If for every vector $\boldsymbol{v}$ in V, $[\boldsymbol{v}]_{\mathcal{B}} = [\boldsymbol{v}]_{\mathcal{B}'}$ then $\mathcal{B}' = \mathcal{B}$.

22. $P_{\mathcal{B} \to \mathcal{B}'} P_{\mathcal{B}' \to \mathcal{B}} = I_3$, the 3×3 **identity matrix**.

23. $[\boldsymbol{v}_2]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

24. If $\mathcal{B}' \neq \mathcal{B}$ then for every vector $\boldsymbol{v}$, $[\boldsymbol{v}]_{\mathcal{B}'} \neq [\boldsymbol{v}]_{\mathcal{B}}$.

Challenge Exercises (Problems)

1. Let $\mathcal{M} = \left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \right)$.

Suppose $[\boldsymbol{m}]_{\mathcal{M}} = \begin{pmatrix} 3 \\ -2 \\ 1 \\ 4 \end{pmatrix}$. Compute $\boldsymbol{m}$. See **Method** (5.4.2).

2. Assume that $\mathcal{F}' = (1-x, 1-x^2, 1-2x+2x^2)$, $\mathcal{F}$ is a **basis** of $\mathbb{R}_2[x]$ and

$P_{\mathcal{F}\to\mathcal{F}'} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 3 \\ 0 & 2 & 1 \end{pmatrix}$. Determine the vectors in $\mathcal{F}$. See **Method** (5.4.2).

3. Assume $\mathcal{B}$ is a basis for $\mathbb{R}^3$ and

$P_{\mathcal{S}\to\mathcal{B}} = \begin{pmatrix} 2 & 2 & 1 \\ 3 & 4 & 2 \\ 2 & 3 & 2 \end{pmatrix}$.

Find the vectors in $\mathcal{B}$. See **Method** (5.4.4).

4. $\mathcal{B}, \mathcal{B}', \mathcal{B}^*$ are three **bases** of $\mathbb{R}^3$. Assume that

$P_{\mathcal{B}\to\mathcal{B}'} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$, $P_{\mathcal{B}'\to\mathcal{B}^*} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$.

If $[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} 2 \\ -4 \\ 7 \end{pmatrix}$ what is $[\boldsymbol{v}]_{\mathcal{B}^*}$? See **Theorem** (5.4.3).

Quiz Solutions

1. The matrix formed from the **3-vectors** corresponding to $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3$ is

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

Using **Gaussian elimination** we obtain the following **echelon form**:

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

Since every column is a **pivot column** the sequence is **linearly independent**. Since $dim(\mathbb{R}_2[x]) = 3$ this is enough to conclude that the sequence is a **basis** of $\mathbb{R}_2[x]$.

Not right, see **Method** (5.2.3) and **Theorem** (5.3.8).

2. This can be done all at once by adjoining to the matrix A of question 1 each of the **3-vectors** corresponding to the **standard basis vectors** of $\mathbb{R}_2[x]$ and then using **Gaussian elimination** to obtain the **reduced echelon form**. The columns on the right of the **reduced echelon form** will give the coefficients for the corresponding **linear combination**. The initial matrix is

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

The **reduced echelon form** is

$$\begin{pmatrix} 1 & 0 & 0 & | & 4 & -7 & 2 \\ 0 & 1 & 0 & | & -2 & 4 & -1 \\ 0 & 0 & 1 & | & 3 & -5 & 1 \end{pmatrix}$$

Thus,

$$1 = 4\boldsymbol{f}_1 - 2\boldsymbol{f}_2 + 3\boldsymbol{f}_3,$$

$$x = -7\boldsymbol{f}_1 + 4\boldsymbol{f}_2 - 5\boldsymbol{f}_3,$$

$$x^2 = 2\boldsymbol{f}_1 - \boldsymbol{f}_2 + \boldsymbol{f}_3.$$

Not right, see **Method** (5.2.1).

3. We associate to each matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ the **4-vector** $\begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{pmatrix}$. Now make the 4×4 matrix whose columns are the 4-vectors associated with $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$. This matrix is as follows:

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

Using **Gaussian elimination** we obtain the following **echelon form**:

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

Thus, every column is a **pivot column** which implies that $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is **linearly independent**. Since $dim(M_{2\times 2}(\mathbb{R})) = 4$ it follows that $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ is a **basis** of $M_{2\times 2}(\mathbb{R})$.

Not right, see **Method** (5.2.3) and **Theorem** (5.3.8).

4. Adjoin the column associated to $\boldsymbol{m}$ to the matrix B in question 3 and use **Gaussian elimination** to obtain the **reduced echelon form**. The required augmented matrix is

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

The **reduced echelon form** of this matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & -5 \\ 0 & 1 & 0 & 0 & | & 9 \\ 0 & 0 & 1 & 0 & | & -11 \\ 0 & 0 & 0 & 1 & | & 2 \end{pmatrix}$$

Thus $\boldsymbol{m} = -5\boldsymbol{m}_1 + 9\boldsymbol{m}_2 - 11\boldsymbol{m}_3 + 2\boldsymbol{m}_4$.

Not right, see **Method** (5.2.1).

5.5. Rank and Nullity of a Matrix

We return to the **vector space** $\mathbb{R}^n$ and matrices. We define the notions of rank and nullity of a matrix and prove a fundamental relationship between these two numbers and the number of columns of the matrix.

Am I Ready for This Material

The following are concepts and procedures essential for understanding the material of this section:

translate of a subset of $\mathbb{R}^n$

echelon form of a matrix

reduced echelon form of a matrix

pivot column of a matrix

pivot position of a matrix

elementary row operations

row equivalence of matrices

subspace of $\mathbb{R}^n$

linearly dependent sequence of vectors from $\mathbb{R}^n$

linearly independent sequence of vectors from $\mathbb{R}^n$

basis of a subspace of $\mathbb{R}^n$

null space of a matrix

column space of a matrix

row space of a matrix

basis of a subspace of $\mathbb{R}^n$

dimension of a subspace of $\mathbb{R}^n$

invertible matrix

inverse of an invertible matrix

Gaussian elimination

Quiz

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

1. Find the **reduced echelon form** of A.

2. Find a basis for the **null space** of $A, null(A)$.

Let $B = \begin{pmatrix} 1 & 2 & 2 & 3 & 2 & 1 \\ 1 & 3 & 1 & 2 & 1 & 4 \\ 1 & 3 & 1 & 3 & 1 & 6 \\ 1 & 3 & 1 & 2 & 1 & 4 \end{pmatrix}$.

3. Determine the **pivot columns** of B.

Quiz Solutions

New Concepts

The important concepts which are introduced in this section are:

nullity of a matrix

column rank of a matrix

row rank of a matrix

rank of a matrix

Theory (Why It Works)

Let us recall a definition we previously made regarding matrices:

Definition (3.8)

For an $m \times n$ matrix A the ***null space of*** A, denoted by null(A), is defined to be the collection of all $n-$vectors $\boldsymbol{v}$ which satisfy $A\boldsymbol{v} = \mathbf{0}_m$.

Since the **null space** of an $m \times n$ matrix is a **subspace** of $\mathbb{R}^n$ it has a **dimension**. Because this dimension has both theoretic and practical interest we give it a name:

Definition 5.18. The **dimension** of the **null space** of a matrix A is referred to as the ***nullity of*** A and is denoted by $nullity(A)$.

In the following we show that **row equivalent matrices** have identical **null spaces**:

Theorem 5.5.1. *Let A and B be $m \times n$ matrices and assume that A and B are* **row equivalent matrices** *that is, A can be transformed into B via* **elementary row operations***. Then the* **null space** *of A is equal to the null space of $B, null(A) = null(B)$.*

Proof. Assume that A and B are **row equivalent**. By **Theorem** (3.5.5) there is an **invertible** $m \times m$ matrix Q such that $QA = B$. Suppose $\boldsymbol{v} \in null(A)$. Then $B\boldsymbol{v} = (QA)\boldsymbol{v} = Q(A\boldsymbol{v}) = Q\mathbf{0}_m = \mathbf{0}_m$ and therefore $\boldsymbol{v} \in null(B)$. This shows that $null(A)$ is contained in $null(B)$.

Now let P be the **inverse** of $Q, P = Q^{-1}$, so that $A = PB$. Then by the same reasoning we get $null(B)$ is contained in $null(A)$ and therefore $null(A)$ and $null(B)$ are equal. $\square$

Example 5.5.1. Let $A = \begin{pmatrix} 1 & -2 & -1 & -1 & -1 \\ 2 & -4 & -1 & -2 & 0 \\ 4 & -8 & -4 & -3 & -6 \\ 1 & -2 & -2 & 0 & -5 \end{pmatrix}$. Find a **basis** for **null(A)** and use this to compute the **nullity** of A.

To find a **basis** for **null(A)** we use **Gaussian elimination** to obtain the **reduced echelon form**. This is the matrix

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

We now write out the **homogeneous linear system** which has the matrix B as its **coefficient matrix**:

$$\begin{array}{ccccccccc} x_1 & - & 2x_2 & & & & - & x_5 & = & 0 \\ & & & x_3 & & & + & 2x_5 & = & 0 \\ & & & & & x_4 & - & 2x_5 & = & 0 \end{array}$$

There are three **leading variables** (x_1, x_3, x_4) and two **free variables** (x_2, x_5). We choose parameters for the free variables $(x_2 = s, x_5 = t)$ and express the **leading variables** in terms of the parameters:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 2s + t \\ s \\ -2t \\ 2t \\ t \end{pmatrix} = s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ -2 \\ 2 \\ 1 \end{pmatrix}.$$

Therefore, $null(A) = Span\left(\begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -2 \\ 2 \\ 1 \end{pmatrix}\right)$. Since the vectors $\begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -2 \\ 2 \\ 1 \end{pmatrix}$ are not multiples of each other, the sequence $\left(\begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -2 \\ 2 \\ 1 \end{pmatrix}\right)$ is **linearly independent** and, consequently, $dim(null(A)) = 2$. Thus, the **nulllity** of A is two. *Notice that this is the number of* **free variables**.

Recall that for an $m \times n$ matrix A we previously defined the **column space** of A, denoted by $col(A)$, to be the **subspace** of $\mathbb{R}^m$ **spanned** by the columns of A. Obviously this has a **dimension** which is the subject of the next definition.

Definition 5.19. For an $m \times n$ m matrix A the ***column rank*** of A is the **dimension** of the **column space** of A. We denote this by $colrank(A)$. Thus, $colrank(A) = dim(col(A))$.

Recall that we pointed out that two $m \times n$ **row equivalent matrices** A and B have identical **null spaces**. This can be interpreted in the following way:

Theorem 5.5.2. *Assume the $m \times n$ matrices A and B are* **row equivalent**. *Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ be the columns of A and $\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_n$ be the columns of B. Then for a sequence of scalars $(c_1, c_2, \dots, c_n)$ $c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + c_n\boldsymbol{a}_n = 0$ if and only if $c_1\boldsymbol{b}_1 + c_2\boldsymbol{b}_2 + \cdots + c_n\boldsymbol{b}_n = 0$.*

Proof. This follows immediately from **Theorem** (5.5.1) and the **definition of the product of a matrix and a vector**: A vector $\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ is in **null(A)** if and only if $A\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = c_1\boldsymbol{a}_1 + \cdots + c_n\boldsymbol{a}_n = \boldsymbol{0}_m$. □

A consequence of **Theorem** (5.5.2) is that a subsequence of the columns of A is **linearly dependent** if and only if the corresponding subsequence of the columns of B is linearly dependent. If B is in **echelon form** then it is transparent to see when a subsequence of the columns is **linearly dependent**: more precisely, a subsequence of columns of a matrix in echelon form is linearly dependent if and only if it contains a **non-pivot column**. In fact, every non-pivot column is a **linear combination** of the **pivot columns** which precede it.

Example 5.5.2. Let A and B be the matrices from **Example** (5.5.1) The first thing to notice is that the **column space** of B is contained in the **subspace** of $\mathbb{R}^4$ consisting of those vectors whose last entry is zero. Consequently, the **column rank** of B is at most three.

On the other hand look at the **pivot columns** of B. These are: $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\0\end{pmatrix}$.

This is a subsequence of the **standard basis** of $\mathbb{R}^4$ and therefore **linearly independent**. Therefore the **column space** of B contains $Span(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ which has **dimension** 3. We have shown that $dim(col(B)) \geq 3$ and $dim(col(B)) \leq 3$ and therefore $dim(col(B)) = 3$.

By **Theorem** (5.5.2) it follows that columns 1, 3 and 4 of A, the **pivot columns**, are a **basis** for the **column space** of A :

$$col(A) = Span\left(\begin{pmatrix}1\\2\\4\\1\end{pmatrix}, \begin{pmatrix}-1\\-1\\-4\\-2\end{pmatrix}, \begin{pmatrix}-1\\-2\\-3\\0\end{pmatrix}\right).$$

Implicit in what we have just done is a method for finding a **basis** for a **subspace** of $\mathbb{R}^n$ when given a **spanning sequence**.

Computation of a basis for Span(S)

Given a sequence $S = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)$ from $\mathbb{R}^m$ in order to find a **basis** for $Span(S)$ do the following:

1. Form the matrix A with the vectors from S as columns.

2. Apply **Gaussian elimination** to obtain an **echelon form**, E, which is used to determine the **pivot columns** of E and A.

3. The sequence of **pivot columns** of A is a **basis** for $Span(S)$.

Remark 5.10. Each **non-pivot column** is a **linear combination** of the **pivot columns** which precede it. Consequently, if a sequence of vectors is **linearly independent** and make up the initial columns of a matrix then each one will be a **pivot column** of that matrix.

Based on **Remark** (5.10) we have the following procedure for extending a **linearly independent** sequence of vectors to a **basis** of $\mathbb{R}^m$.

Extending a linearly independent set to a basis

Suppose $S = (\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k)$ is a **linearly independent** sequence in $\mathbb{R}^m$ (so $k \leq m$). In order to extend the sequence to a **basis** make a matrix B starting with $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ as columns and then follow with the **standard basis** $(\boldsymbol{e}_1^m, \boldsymbol{e}_2^m, \dots, \boldsymbol{e}_m^m)$ of $\mathbb{R}^m$ (actually any basis will work but the standard basis is always accessible).

Since the sequence of columns contains the **standard basis** of $\mathbb{R}^m$ it follows the **column space** of B is $\mathbb{R}^m$. Apply **Gaussian elimination** to find an **echelon form** of B in order to identify the **pivot columns** of B. This will be a **basis** of $\mathbb{R}^m$ which contains the sequence $(\boldsymbol{a}_1, \dots, \boldsymbol{a}_k)$.

Example 5.5.3. Let $\boldsymbol{a}_1 = \begin{pmatrix} 1 \\ -2 \\ 2 \\ -1 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} 2 \\ -3 \\ 4 \\ -2 \end{pmatrix}$. Extend $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ to a **basis** for $\mathbb{R}^4$.

We make the matrix with columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{e}_1^4, \boldsymbol{e}_2^4, \boldsymbol{e}_3^4, \boldsymbol{e}_4^4$, apply **Gaussian elimination** to find an **echelon form**, and identify the **pivot columns**. The relevant matrix and the sequence of **elementary row operations** is as follows:

$$A = \begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 0 \\ -2 & -3 & 0 & 1 & 0 & 0 \\ 2 & 4 & 0 & 0 & 1 & 0 \\ -1 & -2 & 0 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \to$$
$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{pmatrix}. \tag{5.30}$$

The matrix in (5.30) is an **echelon form** of A and its **pivot columns** are 1,2,3,5 and therefore $(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{e}_1^4, \boldsymbol{e}_3^4)$ is a **basis** of $\mathbb{R}^4$ which extends $(\boldsymbol{a}_1, \boldsymbol{a}_2)$.

Recall that for an $m \times n$ matrix A we previously defined the **row space** of A to be the **subspace** of $\mathbb{R}^n$ **spanned** by the rows of A. This is denoted this by $row(A)$.

Definition 5.20. Let A be an $m \times n$ matrix. The ***row rank*** of A is the **dimension** of the **row space**. The row rank of a matrix A is denoted by $rowrank(A)$. Thus, $rowrank(A) = dim(row(A))$.

By **Theorem** (2.3.9) matrices A and B which are **row equivalent** have the same **row space**. In particular, this holds for a matrix A and its **reduced echelon form**. Implicit in this is a method for finding a **basis** of the **row space** of the matrix A.

Finding a basis for the row space of a matrix

Given a matrix A use **Gaussian elimination** to obtain the **reduced echelon form**, E. Then the **row space** of A and E are identical. However, the sequence of non-zero rows of E is **linearly independent** and therefore a **basis** for the **row space** of E and therefore for the row space of A.

We demonstrate why this is so by reference to an example:

Example 5.5.4. Let

$$A = \begin{pmatrix} 1 & -2 & -1 & -1 & -1 \\ 2 & -4 & -1 & -2 & 0 \\ 4 & -8 & -4 & -3 & -6 \\ 1 & -2 & -2 & 0 & -5 \end{pmatrix}.$$

We have seen in **Example** (5.5.1) that the **reduced echelon form** of this matrix is

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

If we multiply the first three rows of B, respectively, by scalars c_1, c_2, c_3 and add up then in positions corresponding to the **pivot columns** we get, respectively, c_1, c_2, c_3. If this is to be zero vector then we must have $c_1 = c_2 = c_3 = 0$. This gives a second way to get a basis for a subspace given as a span of a set of vectors:

Alternative computation of a basis for Span(S).

Given a sequence $S = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k)$ of vectors from $\mathbb{R}^n$, to find a **basis** of Span(S) do the following:

1. Make a matrix with these vectors treated as **rows**.

2. Use **Gaussian elimination** to obtain the **reduced echelon form**..

3. The sequence of the **non-zero rows** of the **reduced echelon form** (treated as columns) is the desired **basis** .

Example 5.5.5. Find a basis for the **column space** of the matrix

$$A = \begin{pmatrix} 1 & -2 & -1 & -1 & -1 \\ 2 & -4 & -1 & -2 & 0 \\ 4 & -8 & -4 & -3 & -6 \\ 1 & -2 & -2 & 0 & -5 \end{pmatrix}.$$

We previously did this in **Example** (5.5.1). Now we use the **row method**. Since we must treat the columns as rows we take the **transpose** of A and then apply **Gaussian elimination** to compute its **reduced echelon form**.

$$A^{Tr} = \begin{pmatrix} 1 & 2 & 4 & 1 \\ -2 & -4 & -8 & -2 \\ -1 & -1 & -4 & -2 \\ -1 & -2 & -3 & 0 \\ -1 & 0 & -6 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 4 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & -2 & -4 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 0 & 4 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Therefore $\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right)$ is a **basis** for the **column space** of A.

At this point we can prove that a matrix has a unique **reduced echelon form**, as promised back in Section (1.2).

Theorem 5.5.3. **Uniqueness of the reduced echelon form**

If A be an $m \times n$ matrix then the **reduced echelon form** *of A is unique.*

Proof. Let E and E' be two matrices which are in **reduced row echelon form** and which are both **row equivalent** to A. Let the sequence of **non-zero rows** of E be $(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_k)$ and the sequence of non-zero rows of E' be $(\boldsymbol{r}'_1, \boldsymbol{r}'_2, \dots, \boldsymbol{r}'_k)$ (k is the number of pivots). Now E and E' have the same **pivot columns** by **Theorem** (5.5.2) which is a subsequence of the **standard basis** of $\mathbb{R}^m$. This implies that when $\boldsymbol{r}'_j$ is written as **linear combination** of the $(\boldsymbol{r}_1, \dots, \boldsymbol{r}_k)$ we must have all coefficients equal to zero except the j^{th}, which is a one. Thus, $\boldsymbol{r}'_j = \boldsymbol{r}_j$. □

Notice that the **row rank** of a matrix A and the **column rank** of A are the same since the latter is the number of **pivot columns** and the former is the number of **pivot positions** and these are equal. This motivates the following:

Definition 5.21. The common number $dim(row(A)) = dim(col(A))$ is referred to as the ***rank*** of the matrix A.

Remark 5.11. The **row space** of $A, row(A)$, is essentially the same as the **column space** of A^{Tr}, $col(A^{Tr})$. Consequently we have $rank(A) = rank(A^{Tr})$.

We now come to one of the most important theorems of elementary linear algebra. It goes by the name of the "Rank and Nullity Theorem For Matrices".

Theorem 5.5.4. *Let A be a matrix. Then the sum of the* **rank** *of A and the* **nullity** *of A is equal to the number of columns of A. Thus, if A is an $m \times n$ matrix then $rank(A) + nullity(A) = n$.*

Proof. The **rank** of A is the number of **pivot columns** or, what amounts to the same thing, the number of **leading variables** in the **homogeneous linear system** which has A as its **coefficient matrix**. The **nullity** of A is the number of **non-pivot columns**, equivalently, the number of **free variables** of the aforementioned homogeneous linear system. $\square$

As a corollary of the **Rank-Nullity Theorem** we get the following result relating the **dimension** of a **subspace** W of $\mathbb{R}^n$ and the **dimension** of its **orthogonal complement**, $W^\perp = \{\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{v} \cdot \boldsymbol{w} = 0 \text{ for all } \boldsymbol{w} \in W\}$.

Theorem 5.5.5. *Let W be a* **subspace** *of $\mathbb{R}^n$. Then $dim(W) + dim(W^\perp) = n$.*

Proof. Let $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ be a **basis** for W, a **subspace** of $\mathbb{R}^n$. Let A be the **transpose** of the matrix with $\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k$ as columns: $A = (\boldsymbol{w}_1\boldsymbol{w}_2 \dots \boldsymbol{w}_k)^{Tr}$. Now $rank(A) = dim(row(A)) = k = dim(W)$. By the rank and nullity theorem, $rank(A) + nullity(A) = n$. However, $null(A) = W^\perp$ from which the theorem now follows. $\square$

The next several theorems will fulfill the promise made in Section (1.2), specifically we will show that two **equivalent linear systems** L and L' of m equations in n variables have **equivalent** augmented matrices.

Theorem 5.5.6. *Let A and B be $m \times n$ matrices and assume that $row(A) = row(B)$. Then A and B are* **row equivalent***.*

Proof. Assume the **rank** of A is k. Then some subsequence of the rows of A with length k is **linearly independent**. By performing a sequence of exchange operations we can obtain a matrix A' (which is equivalent to A) such that the sequence of the first k rows is **linearly independent**. Likewise we can find a matrix B' which is **row equivalent** to B such that the sequence of the first k rows is **linearly independent**. If A' is **row equivalent** to B' then A and B are equivalent and consequently, without any loss of generality, we may assume the sequence of the first k rows of A is **linearly independent** and therefore a **basis** for $row(A)$ and similarly for B.

Let $(\boldsymbol{r}_1, \dots, \boldsymbol{r}_m)$ be the sequence of rows of A. Since each $\boldsymbol{r}_l, l > k$, is a **linear combination** of $(\boldsymbol{r}_1, \dots, \boldsymbol{r}_k)$ we can apply a sequence of elimination type **elementary row operations** to A to obtain a matrix A^* with rows $(\boldsymbol{r}_1, \dots, \boldsymbol{r}_k, \mathbf{0}_n^{Tr}, \dots, \mathbf{0}_n^{Tr})$. Then A and A^* are **row equivalent**. Likewise if the sequence of rows in B is $(\boldsymbol{r}'_1, \dots, \boldsymbol{r}'_m)$ then there is a matrix B^* which is **row equivalent** to B with rows $(\boldsymbol{r}'_1, \dots, \boldsymbol{r}'_k, \mathbf{0}_n^{Tr}, \dots, \mathbf{0}_n^{Tr})$.

If A^* and B^* are **row equivalent** then so are A and B so we may as well assume that $A = A^*$ and $B = B^*$, that is, the sequence of the first k rows of A is a basis for

$row(A)$ and all other rows are zero and likewise for B. In this case, A and B will be **row equivalent** if and only if the matrix consisting of the non-zero rows of A is row equivalent to the matrix consisting of the non-zero rows of B. So, without any loss of generality we may assume that $k = m$.

Now let E_A be the **reduced echelon form** of A and E_B be the reduced echelon form of B. We will show that $E_A = E_B$ which will complete the proof. Since we are assuming that the **rank** of A is m, it follows that there is an $m \times (n-m)$ matrix C such that $E_A = (I_m\ C)$ and similarly an $m \times (n-m)$ matrix D such that $E_B = (I_m\ D)$. (Note, in general, if X is an $a \times b$ matrix and Y is an $a \times c$ matrix then by $(X\ Y)$ we mean the $a \times (b+c)$ matrix whose first b columns coincide with the columns of X followed by the columns of Y).

Since A and E_A are **row equivalent** it follows by **Theorem** (2.3.9) that $row(E_A) = row(A)$ and likewise $row(E_B) = row(B)$. Thus, $row(E_A) = row(E_B)$. This implies that each row of E_B is a **linear combination** of the rows of E_A. For $1 \leq i \leq m$ let $\boldsymbol{f}_i = (\boldsymbol{e}_i^m)^{Tr}$, let $\boldsymbol{c}_i$ be the i^{th} row of C and $\boldsymbol{d}_i$ the i^{th} row of D. Then the i^{th} row of E_A is $(\boldsymbol{f}_i\ \boldsymbol{c}_i)$ and the i^{th} row of E_B is $(\boldsymbol{f}_i\ \boldsymbol{d}_i)$. Since each row of E_B is a **linear combination** of the rows of E_A it follows that there are scalars a_{ij} such that

$$\begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & \boldsymbol{d}_j \end{pmatrix} = (\boldsymbol{f}_j\ \boldsymbol{d}_j) =$$

$$a_{1j}(\boldsymbol{f}_1\ \boldsymbol{c}_1) + \dots + a_{mj}(\boldsymbol{f}_m\ \boldsymbol{c}_m) = \begin{pmatrix} a_{1j} & \dots & a_{mj} & [a_{1j}\boldsymbol{c}_1 + \dots + a_{mj}\boldsymbol{c}_m] \end{pmatrix}$$

However, this implies that $a_{jj} = 1$ and $a_{ij} = 0$ for $i \neq j$ and therefore $\boldsymbol{c}_j = \boldsymbol{d}_j$ and the rows of E_A and E_B are identical. $\square$

Theorem 5.5.7. *Let A and B be $m \times n$ matrices and assume that the **null space** of A and B are identical. Then A and B are **row equivalent**.*

Proof. Let E_A be the **reduced echelon form** of A and E_B the **reduced echelon form** of B. By **Theorem** (5.5.1) the **null space** of A and E_A are equal and the null space of B and E_B are equal. Therefore, the **null spaces** of E_A and E_B are equal. It therefore suffices to prove that $E_A = E_B$. As in the proof of **Theorem** (5.5.6) there is no loss of generality in assuming that A and B, and hence E_A and E_B, have rank m. In this case there are $m \times (n-m)$ matrices C and D such that $E_A = (I_m\ C)$ and $E_B = (I_m\ D)$ where I_m is the $m \times m$ **identity matrix**. We need to prove that $C = D$. Let the j^{th} column of C be $\boldsymbol{c}_j = \begin{pmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{mj} \end{pmatrix}$ and the j^{th} column of D be $\boldsymbol{d}_j = \begin{pmatrix} d_{1j} \\ d_{2j} \\ \vdots \\ d_{mj} \end{pmatrix}$.

Now consider the **n-vector** $\boldsymbol{x}_j = \begin{pmatrix} -\boldsymbol{c}_j \\ \boldsymbol{e}_j^{n-m} \end{pmatrix}$. That is, in the first m components $\boldsymbol{x}_j$ is identical with $-\boldsymbol{c}_j$ and then in the next $n-m$ components it is identical with the j^{th} **standard basis vector**. Then $\boldsymbol{x}_j$ is in the **null space** of E_A. Since $null(E_A) = null(E_B)$ it follows that $\boldsymbol{x}_j \in null(E_B)$. On the other hand, the product of $\boldsymbol{x}_j$ with the i^{th} row of E_B is $-c_{ij} + d_{ij}$ and therefore $c_{ij} = d_{ij}$ Since i and j are arbitrary we conclude that $C = D$ as desired. □

We now prove our theorem about **equivalent linear systems**.

Theorem 5.5.8. *Let L be a* **linear system** *with* **augmented matrix** $(A|\boldsymbol{b})$ *and L' a linear system with augmented matrix $(A'|\boldsymbol{b}')$. Assume L and L' are* **equivalent linear systems**. *If A and A' are both $m \times n$ matrices then the matrices $(A|\boldsymbol{b})$ and $(A'|\boldsymbol{b}')$ are* **row equivalent**.

Proof. Recall that L and L' are **equivalent** means that they have the same **solution set**. By **Theorem** (3.2.5) the **solution set** to the **linear system** L can be expressed as $\boldsymbol{p} + null(A)$ where $A\boldsymbol{p} = \boldsymbol{b}$. Since this is also the solution set to L' we can conclude that the **null space** of A' is equal to the null space of $A, null(A') = null(A)$. By **Theorem** (5.5.7) the matrices A and A' are **row equivalent**. By **Theorem** (3.5.5) there is an **invertible** matrix Q such that $QA = A'$. Then the matrix $Q(A|\boldsymbol{b}) = (QA|Q\boldsymbol{b}) = (A'|Q\boldsymbol{b})$ is **row equivalent** to $(A|\boldsymbol{b})$. Let L^* be the **linear system** with **augmented matrix** $(A'|Q\boldsymbol{b})$. Since the matrices $(A|\boldsymbol{b})$ and $(A'|Q\boldsymbol{b})$ are **row equivalent** the linear systems L and L^* are **equivalent**. But this implies that the **solution set** of L^* is $\boldsymbol{p} + null(A)$ which means that $A'\boldsymbol{p} = Q\boldsymbol{b}$. However, since $\boldsymbol{p} + null(A)$ is the solution set to L' it follows that $A'\boldsymbol{p} = \boldsymbol{b}'$. Thus $Q\boldsymbol{b} = \boldsymbol{b}'$ and therefore $(A|\boldsymbol{b})$ and $(A'|\boldsymbol{b}')$ are **row equivalent**. □

What You Can Now Do

1. Given a finite sequence S of vectors from $\mathbb{R}^n$ find a subsequence S_0 of S which is a **basis** for $Span(S)$.

2. Given a matrix A find a **basis** for the **row space** of A.

3. Given a finite sequence S of **n-vectors** find a **basis** for $Span(S)$ for which the computation of **coordinate vectors** is simple.

4. Given a **linearly independent** sequence of **n-vectors** , $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$, extend S to a **basis** of $\mathbb{R}^n$.

Method (How To Do It)

Method 5.5.1. Given a finite sequence S of vectors from $\mathbb{R}^n$ find a subsequence S_0 of S which is a **basis** for $Span(S)$.

If $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ is a sequence from $\mathbb{R}^n$ make the matrix A with these vectors as its columns. Use **Gaussian elimination** to obtain any **echelon form** and use it to determine the **pivot columns**. The original vectors corresponding to the **pivot columns** is a **basis** for $Span(S)$.

Example 5.5.6. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 2 \\ 3 \\ 2 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 3 \\ 1 \\ 3 \\ 1 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 1 \end{pmatrix},$ $\boldsymbol{v}_5 = \begin{pmatrix} 3 \\ 4 \\ 3 \\ 6 \\ 3 \end{pmatrix}, \boldsymbol{v}_6 = \begin{pmatrix} 3 \\ 5 \\ 3 \\ 4 \\ 2 \end{pmatrix}$. Find a **basis** for $Span\,(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_6)$.

Make the matrix A with these vectors as columns:

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

The **reduced echelon form** for A, obtained by **Gaussian elimination** , is

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

The **pivot columns** are 1, 2, 4 and 6. Therefore $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4, \boldsymbol{v}_6)$ is **basis** for the **span** of $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_6)$.

Method 5.5.2. Given a matrix A compute a **basis** for the **row space** of A.

Use **Gaussian elimination** to obtain the **reduced echelon form** of A. The sequence of **non-zero rows** of the **reduced echelon form** is a **basis** for the **row space** of the matrix A.

Example 5.5.7. Find a **basis** for the **row space** of the matrix $A =$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 2 & 2 & 2 \\ 3 & 5 & 1 & 3 & 3 & 3 \\ 3 & 4 & 2 & 2 & -1 & 2 \\ 1 & 2 & 0 & 2 & 5 & 2 \end{pmatrix}.$$

By **Gaussian elimination** we obtain the **reduced echelon form**:

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

The sequence $\left(\begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \\ -3 \\ 0 \end{pmatrix}^{Tr}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}^{Tr}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 4 \\ 1 \end{pmatrix}^{Tr} \right)$ is **basis** for the **row space** of A.

Method 5.5.3. Given a finite sequence S of **n-vectors** find a **basis** for $Span(S)$ for which the computation of **coordinate vectors** is simple.

Make the matrix A with the vectors of S as its columns. Apply **Gaussian elimination** to the **transpose** of A to obtain the **reduced echelon form** U of A^{Tr}. Let $\mathcal{B}$ be the sequence of non-zero columns of U^{Tr}. Then $\mathcal{B}$ is the desired **basis**.

Example 5.5.8. We use the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_6$ of **Example** (5.5.6).

Note that the matrix we apply **Gaussian elimination** to is the **transpose** of the matrix of that example.

The matrix with these vectors as rows is

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

By application of **Gaussian elimination** we obtain the **reduced echelon form**:

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

Set $\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{u}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$. Then $\mathcal{B} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4)$ is a **basis** for $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_6)$.

The **coordinate vector** for a vector $\boldsymbol{v}$ in $Span(S)$ is easy to compute with respect to $\mathcal{B}$. Namely the components of the **coordinate vector** are the components of $\boldsymbol{v}$ that occur in places 1,2,4,5 (the indices of the **pivot columns**). For the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_6)$ from **Example** (5.5.6) the **coordinate vectors** are computed below:

$$\boldsymbol{v}_1 = \boldsymbol{u}_1 + 2\boldsymbol{u}_2 + 2\boldsymbol{u}_3 + \boldsymbol{u}_4, [\boldsymbol{v}_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix},$$

$$\boldsymbol{v}_2 = 2\boldsymbol{u}_1 + 3\boldsymbol{u}_2 + 3\boldsymbol{u}_3 + 2\boldsymbol{u}_4, [\boldsymbol{v}_2]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 2 \end{pmatrix}$$

$$\boldsymbol{v}_3 = \boldsymbol{u}_1 + 3\boldsymbol{u}_2 + 3\boldsymbol{u}_3 + \boldsymbol{u}_4, [\boldsymbol{v}_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 3 \\ 3 \\ 1 \end{pmatrix},$$

$$\boldsymbol{v}_4 = \boldsymbol{u}_1 + \boldsymbol{u}_2 + 2\boldsymbol{u}_3 + \boldsymbol{u}_4, [\boldsymbol{v}_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\boldsymbol{v}_5 = 3\boldsymbol{u}_1 + 4\boldsymbol{u}_2 + 6\boldsymbol{u}_3 + 3\boldsymbol{u}_4, [\boldsymbol{v}_1]_{\mathcal{B}} = \begin{pmatrix} 3 \\ 4 \\ 6 \\ 3 \end{pmatrix},$$

$$\boldsymbol{v}_6 = 3\boldsymbol{u}_1 + 5\boldsymbol{u}_2 + 4\boldsymbol{u}_3 + 2\boldsymbol{u}_4, [\boldsymbol{v}_1]_{\mathcal{B}} = \begin{pmatrix} 3 \\ 5 \\ 4 \\ 2 \end{pmatrix}$$

If $\mathcal{B}' = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4, \boldsymbol{v}_6)$, which we orginally obtained as a **basis** of $Span(S)$ in **Example** (5.5.6), then the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$ is $\begin{pmatrix} 1 & 2 & 1 & 3 \\ 2 & 3 & 1 & 5 \\ 2 & 3 & 2 & 4 \\ 1 & 2 & 1 & 2 \end{pmatrix}$.

Method 5.5.4. Given an independent sequence of **n-vectors** , $S = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$, extend S to a **basis** of $\mathbb{R}^n$.

Make the matrix A whose first k columns are the vectors of the sequence S followed by the **standard basis vectors**, $\boldsymbol{e}_1^n, \boldsymbol{e}_2^n, \dots, \boldsymbol{e}_n^n$. Use **Gaussian elimination** to obtain an **echelon form** and determine the **pivot columns**. The first k **pivot columns** will be the vectors of the sequence S. The **pivot columns** of A will be a **basis** of $\mathbb{R}^n$ which extends S.

Example 5.5.9. Extend $S = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ -1 \\ 2 \\ -1 \end{pmatrix} \right)$ to a **basis** of $\mathbb{R}^5$.

Make the matrix with the three vectors as the first three columns and complete it with $\boldsymbol{e}_1^5, \boldsymbol{e}_2^5, \dots, \boldsymbol{e}_5^5$. This matrix follows:

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

We use **Gaussian elimination** to obtain an **echelon form**:

$$\begin{pmatrix} 1 & 1 & 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 & -\frac{7}{3} & \frac{1}{3} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}.$$

The **pivot columns** are 1,2,3,4,6 and the **basis** $\mathcal{B}$ extending S is

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

Exercises

In exercises 1-4 use **Method** (5.5.1) to find a **basis** for the **subspace** of $\mathbb{R}^n$ which is **spanned** by the given sequence of vectors.

1. $\left(\begin{pmatrix} 1 \\ 2 \\ -3 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 7 \\ -5 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -7 \\ -2 \end{pmatrix} \right)$

2. $\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \right)$

3. $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ -3 \\ -3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \\ 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 2 \\ 1 \\ 2 \end{pmatrix} \right)$

4. $\left(\begin{pmatrix}1\\1\\-2\\1\\0\\-1\end{pmatrix}, \begin{pmatrix}2\\3\\-5\\2\\1\\-3\end{pmatrix}, \begin{pmatrix}1\\2\\-3\\1\\1\\-2\end{pmatrix}, \begin{pmatrix}0\\1\\-1\\0\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\\0\\1\\2\\-3\end{pmatrix}, \begin{pmatrix}3\\2\\-5\\3\\1\\-4\end{pmatrix}\right)$

In exercises 5-8 find a **basis** for the **row space** of the given matrix. See **Method** (5.5.2).

5. $\begin{pmatrix}2&3&3&2&2\\3&4&4&2&2\\1&2&2&2&2\\1&1&1&1&1\\0&1&1&2&2\end{pmatrix}$

6. $\begin{pmatrix}1&1&2&2&3&3\\2&2&4&4&5&5\\0&0&0&0&1&1\\1&1&2&2&2&2\\1&2&1&2&1&2\end{pmatrix}$

7. $\begin{pmatrix}2&5&1&2\\1&3&2&5\\1&4&4&9\\1&2&-1&-3\\1&2&-1&1\end{pmatrix}$

8. $\begin{pmatrix}1&1&-2&2&-4&2\\1&2&-3&2&-5&3\\2&1&-3&4&-7&3\\0&1&-1&0&-1&1\\1&3&-4&-3&-1&4\end{pmatrix}$

In exercises 9 - 12: a) Find a **basis** $\mathcal{B}$ for the **span** for the given sequence S of vectors that is efficient for the purposes of finding **coordinate vectors**; and b) For each vector $\boldsymbol{v}$ in S compute the **coordinate vector** of $\boldsymbol{v}$ with respect to $\mathcal{B}$. See **Method** (5.5.3).

9. S is the sequence of vectors of exercise 1.

10. S is the sequence of vectors of exercise 2.

11. S is the sequence of vectors of exercise 3.

12. S is the sequence of vectors of exercise 4.

In 13 - 16 extend the given **linearly independent** sequence to a **basis** of $\mathbb{R}^n$.
See **Method** (5.5.4).

13. $\left(\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\2\\1\\2\end{pmatrix}\right)$

14. $\left(\begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix}, \begin{pmatrix}1\\4\\5\\7\\9\end{pmatrix}\right)$

15. $\left(\begin{pmatrix}1\\1\\2\\1\\2\\3\end{pmatrix}, \begin{pmatrix}1\\2\\3\\2\\2\\4\end{pmatrix}, \begin{pmatrix}2\\2\\4\\1\\1\\2\end{pmatrix}, \begin{pmatrix}1\\3\\4\\3\\1\\4\end{pmatrix}\right)$

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

In exercises 17 - 24 answer true or false and give an explanation.

17. If A is a 5×5 matrix then the zero vector is the only vector that **null(A)** and **col(A)** have in common.

18. If $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ is a sequence of **linearly independent** vectors in $\mathbb{R}^5$ then $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{e}_1^5, \boldsymbol{e}_2^5)$ is a **basis** of $\mathbb{R}^5$ where $\boldsymbol{e}_1^5 = \begin{pmatrix}1\\0\\0\\0\\0\end{pmatrix}$ and $\boldsymbol{e}_2^5 = \begin{pmatrix}0\\1\\0\\0\\0\end{pmatrix}$.

19. If A is a 4×4 matrix and $A^2 = \mathbf{0}_{4\times 4}$ then **nullity(A)** is at least two.

20. If A and B are matrices with $null(A) = null(B)$ then A and B are **row equivalent**.

21. If A and B are $m \times n$ matrices and $nullity(A) = nullity(B)$ then A and B are **row equivalent**.

22. If A and B are **row equivalent** matrices then $null(A) = null(B)$.

23. If A is an $m \times n$ matrix and R is the **reduced echelon form** form of A then $col(A) = col(R)$.

24. If A is an $m \times n$ matrix and R is the **reduced echelon form** of A then $row(A) = row(R)$.

Challenge Exercises (Problems)

1. Let A be a 7×9 matrix.

a) Assume that the **rank** of A is 8. What is the **nullity** of A?

b) Assume that the **rank** of A is 5. What is the **nullity** of A?

c) Assume that the **rank** of A^{Tr} is 5. What is the **nullity** of A?

d) Assume that the **nullity** of A is 3. What is the **rank** of A^{Tr}?

e) Assume that the **nullity** of A^{Tr} is 3. What is the **nullity** of A?.

See **Theorem** (5.5.4).

2. Let A be a 6×8 matrix.

a) Suppose A has 2 **non-pivot columns**. Explain why the **pivot columns** of A are a **basis** of $\mathbb{R}^8$.

b) Suppose A has 4 **non-pivot columns**. What is the **dimension** of the **column space** of A?

c) Suppose $dim(null(A)) = 3$. What is the $dim(null(A^{Tr}))$?

d) Suppose A has 2 **non-pivot columns**. Explain why the **pivot columns** of A are a **basis** of $\mathbb{R}^6$.

e) Suppose $dim(row(A)) = 5$. What is the **nullity** of the **transpose** of A.

See **Theorem** (5.5.4).

3. Let A be an $m \times n$ matrix. Prove the following are equivalent:
a) The **rank** of A is m.

b) The **column space** of A is $\mathbb{R}^m$.

c) A has a **right inverse**, that is, there is an $n \times m$ matrix B such that $AB = I_m$, the $m \times m$ **identity matrix** .

4. Let A be an $m \times n$ matrix. Prove the following are equivalent:

a) The **rank** of A is n;

b) The sequence of columns of A **linearly independent**;
c) A has a **left inverse**, that is, a $n \times m$ matrix C such that $CA = I_n$, the $n \times n$ **identity matrix** .

Quiz Solutions

1. The **reduced echelon form** of A is

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

Not right, see **Method** (1.2.3).

2. $null(A) = Span\left(\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix},\right).$

Not right, see **Method** (1.2.4)..

3. B has an **echelon form** equal to

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

There are three **pivot columns** which are 1, 2 and 4.

Not right, see **Method** (1.2.3).

5.6. Complex Vector Spaces

In this section we review the concept of a **field** and indicate how a **vector space** can be defined with scalars other than the real numbers, $\mathbb{R}$, in particular, the complex numbers, $\mathbb{C}$.

Am I Ready for This Material

In this section we define the notion of a **vector space** over an arbitrary **field**. Since all the concepts originally introduced for real **vector spaces** apply to this more general concept it is very important that these definitions have been mastered. The most relevant ones to the discussion in this section are as follows:

field of complex numbers

$\mathbb{R}^n$

vector space

linear combination of a sequence (or set) of vectors from a vector space

subspace of a vector space

spanning sequence of a vector space

linearly dependent sequence of vectors of a vector space

linearly independent sequence of vectors of a vector space

basis of a vector space

finitely generated vector space

dimension of a finitely generated vector space

finite dimensional vector space

Quiz

1. Compute the sum: $(3-2i)+(-5+7i)$

2. Compute the sum: $(-8+3i)+(-3-5i)$

3. Compute the difference: $(3-4i)-(4-4i)$

4. Compute the difference: $(-2-5i)-(4-7i)$

5. Compute the product: $(3-i)(3+i)$

6. Compute the product: $(1+i)(1+i)$

7. Compute the product: $(3+2i)(4-i)$

Quiz Solutions

New Concepts

A number of new concepts are introduced in this section which will enable us to "extend" the scalars of a real **vector space** to the **complex numbers** and also to define vector spaces over other number systems, in particular, **finite fields**. The former (a complex vector space) is used when we study eigenvectors and eigenvalues in chapter 7 and the latter (a vector space over a finite field) in important applications such as **error correcting codes**. In the present section we will define such notions as **linear independence**, **span**, **basis** and **dimension**, defined previously for a (real) **vector space** , for a complex vector space. The new concepts are that are encountered here are:

field

complex field

vector space over a field

complex vector space

Theory (Why It Works)

In the very first section we briefly analyzed what properties of the (real) numbers are used in solving a single **linear equation** in a single variable. We informally defined a system satisfying these properties as a **field**. Because this concept is at the center of this and the following section we formally write down a rigorous definition.

Definition 5.22. A ***field*** is a set $\mathbb{F}$ which contains two special and distinct elements 0 and 1. It also is equipped with operations $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ called ***addition*** which takes a pair (a, b) in $\mathbb{F}$ to an element $a + b$ in $\mathbb{F}$ and an operation $\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ called ***multiplication*** which takes a pair (a, b) in $\mathbb{F}$ to an element $a \cdot b$. Additionally, $(\mathbb{F}, 0, 1, +, \cdot)$ must satisfy the following axioms:

(A1) For every pair of elements (a, b) from $\mathbb{F}$, $a + b = b + a$. We say that **addition is commutative.**

(A2) For every triple of elements (a, b, c) from $\mathbb{F}$, $a + (b + c) = (a + b) + c$. We say that **addition is associative**

(A3) For every element $a \in \mathbb{F}, a + 0 = a$. We say that **zero is an additive identity**

(A4) Every element a in $\mathbb{F}$ has a **negative** b such that $a + b = 0$. We say that every element has an **additive inverse**.

(M1) For every pair of elements (a, b) from $\mathbb{F}$, $a \cdot b = b \cdot a$. We say that **multiplication is commutative**.

(M2) For every triple of elements (a, b, c) from $\mathbb{F}$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. **Multiplication is associative**.

(M3) For every $a \in \mathbb{F}, a \cdot 1 = a$. **The element 1 is a multiplicative identity**.

(M4) For every $a \in \mathbb{F}, a \neq 0$ there is an element c such that $a \cdot c = 1$. We say that every nonzero element has a **multiplicative inverse**.

(M5) For every triple of elements (a, b, c) from $\mathbb{F}$, $a \cdot (b + c) = a \cdot b + a \cdot c$. This is the **distributive axiom**.

In exactly the same way that we proved in **Theorem** (5.1.1) that the **zero vector** in a **vector space** is unique we can show that the elements 0 and 1 are unique and consequently they are henceforth **the** additive identity and **the** multiplicative identity, respectively. Also we can show the **negative** of an element a of a **field** is unique and we denote this by $-a$. Finally, in a similar way we can show that multiplicative inverses are unique and we will denote the inverse of an element a by a^{-1} or sometimes $\frac{1}{a}$.

Example 5.6.1. The **rational field**, which consists of all numbers which can be expressed as $\frac{m}{n}$ where m, n are integers and $n \neq 0$ is a **field**.

The ***real field*** which consists of all the decimals is also a **field**. Up to this point we have been operating with these examples throughout the book.

Definition 5.23. The ***complex field***, denoted by $\mathbb{C}$, consists of all expresions of the form $a + bi$ where a, b are **real numbers**.

For a complex number $a + bi$ with $a, b \in \mathbb{R}$, a will be called the **real part** and bi the **imaginary part**. When the real part is zero we will write bi for $0 + bi$ and call such a number **purely imaginary**.

We add two complex as follows:

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

Multiplication is defined by

$$(a + bi)(c + di) = (ab - bd) + (ad + bc)i.$$

We will identify a real number a with the complex number $a + 0i$ and in this way the real field, $\mathbb{R}$, is a subset of $\mathbb{C}$.

Remark 5.12. Notice that by the definition of complex multiplication $i^2 = i \cdot i = (0 + i) \cdot (0 + i) = -1 + 0i = -1$. Thus, the pure imaginary number i is a square root of -1.

Definition 5.24. For a complex number $z = a + bi$ $(a, b \in \mathbb{R})$ the ***norm*** of z is defined as $\parallel z \parallel = \sqrt{a^2 + b^2}$.

The ***conjugate*** of $z = a + bi$ is the complex number $\overline{z} = a - bi$.

Remark 5.13. For a complex number z, $\parallel z \parallel = \parallel \overline{z} \parallel$.

Theorem 5.6.1. *1) If z, w are complex numbers then $\parallel zw \parallel \;=\; \parallel z \parallel \times \parallel w \parallel$.*

2) If z is a complex number and c is a real number then $\parallel cz \parallel \;=\; |c| \times \parallel z \parallel$.

3) If $z = a + bi$ is a complex number with $a, b \in \mathbb{R}$ then $z\overline{z} = a^2 + b^2 \;=\; \parallel z \parallel^2$.

Proof. 1) Let $z = a+bi, w = c+di$ with $a, b, c, d \in \mathbb{R}$. Then $zw = (a+bi)(c+di) = (ac - bd) + (ad + bc)i$ and

$$\| zw \|^2 = \| (ac-bd)+(ad+bc)i \|^2 =$$

$$(ac-bd)^2+(ad+bc)^2 =$$

$$(ac)^2 - 2(ac)(bd) + (bd)^2 + (ad)^2 + 2(ad)(bc) + (bc)^2 =$$

$$(ac)^2 + (bd)^2 + (ad)^2 + (bc)^2$$

the last equality since $(ac)(bd) = (ad)(bc) = abcd$. On the other hand

$$(\| z \| \times \| w \|)^2 = \| z \|^2 \times \| w \|^2 =$$

$$(a^2+b^2)(c^2+d^2) =$$

$$a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 =$$

$$(ac)^2 + (ad)^2 + (bc)^2 + (bd)^2$$

We therefore see that $\| zw \|^2 = \| z \|^2 \times \| w \|^2$ and 1) follows by taking square roots.

2) This follows from 1) and the fact that for a real number c, $\| c \| = |c|$.

3) $(a+bi)(a-bi) = [(a)(a)-(b)(-b)] + [(a)(-b)+(b)(a)]i =$

$$a^2+b^2 = \quad \| a+bi \|^2 .$$

□

For later application we require one more theorem about the **complex numbers**, this time dealing with properties of the **conjugate of a complex number**.

Theorem 5.6.2. *1) If z and w are **complex numbers** then $\overline{z+w} = \overline{z} + \overline{w}$.*

*2) If z and w are **complex numbers** then $\overline{zw} = \overline{z} \times \overline{w}$.*

*3) Let z be a **complex number** and c a real number. Then $\overline{cz} = c\overline{z}$.*

Proof. We prove part 2) and leave parts 1) and 2) as **challenge exercises**.

2) Let $z = a + bi, w = c + di$ with a, b, c, d real numbers. Then

$$zw = (ac - bd) + (ad + bc)i, \overline{zw} = (ac - bd) - (ad + bc)i.$$

On the other hand,

$$\overline{z} = a - bi, \overline{w} = c - di$$

$$\overline{z} \times \overline{w} = (a - bi)(c - di) =$$

$$[ac - (-b)(-d)] + [(a)(-d) + (-b)(c)]i =$$

$$(ac - bd) + [-ad - bc]i =$$

$$(ac - bd) - (ad + bc)i$$

Consequently, $\overline{zw} = \overline{z}\,\overline{w}$ as claimed. $\square$

We now demonstrate that the set of **complex numbers** is a **field** which extends the real numbers.

Theorem 5.6.3. *The collection* $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$, *of* **complex numbers**, *with the operations of* **addition and multiplication** *as defined is a field with additive identity* $0 = 0 + 0i$ *and multiplicative identity* $1 = 1 + 0i$.

Proof. (A1) and (A2) hold because real addition is **commutative and associative**. Also, $(a + bi) + 0 = (a + 0) + bi = a + bi$ since 0 is the **additive identity** for $\mathbb{R}$. Thus, (A3) holds with 0.

For (A4) see that

$$[a + bi] + [(-a) + (-b)i] = [a + (-a)] + [b + (-b)]i =$$

$$0 + 0i = 0.$$

Thus, $(-a) + (-b)i$ is an **additive inverse** for $a + bi$. This takes care of the additive axioms.

(M1) $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$ by the definition of multiplication.

$(c+di)(a+bi) = (ca-db)+(cb+da)i$. But $ac = ca, bd = db, ad = da, bc = cb$ since real multiplication is commutative and therefore $(a + bi)(c + di) = (c + di)(a + bi)$ and so complex multiplication is **commutative**.

(M2) We next show that complex multiplication is **associative**. Thus, let $a + bi, c + di, e + fi$ be complex numbers with $a, b, c, d, e, f \in \mathbb{R}$. We need to show that

$$(a + bi)[(c + di)(e + fi)] = [(a + bi)(c + di)](e + fi) \tag{5.31}$$

By the definition of **complex multiplication**

$$(c+di)(e+fi) = (ce-df)+(cf+de)i, (a+bi)(c+di) = (ac-bd)+(ad+bc)i \tag{5.32}$$

It then follows from (5.32) and the definition of **complex multiplication** that

$$(a + bi)[(c + di)(e + fi) = (a + bi)[(ce - df) + (cf + de)i] =$$

$$[a(ce - df) - b(cf + de)] + [a(cf + de) + b(ce - df)]i =$$

$$[ace - adf - bcf - bde] + [acf + ade + bce - bdf]i \tag{5.33}$$

$$[(a + bi)(c + di)](e + fi) = [(ac - bd) + (ad + bc)i](e + fi) =$$

$$[(ac - bd)e - (ad + bc)f] + [(ac - bd)f + (ad + bc)e]i =$$

$$[ace - bde - adf - bcf] + [acf - bdf + ade + bce]i \tag{5.34}$$

Comparing (5.33) and (5.34) we conclude that

$$(a + bi)[(c + di)(e + fi)] = [(a + bi)(c + di)](e + fi).$$

(M3) Let $a + bi$ be a complex number. Then $(a + bi)1 = (a + bi)(1 + 0i) = (a1 - b0) + (b1 + a0)i = a + bi$ and so 1 is a **multiplicative identity**.

(M4) If $z = a + bi \neq 0$ then

$$(z)(\frac{\overline{z}}{\| z \|^2}) = \frac{z\overline{z}}{\| z \|^2} =$$

$$\frac{\| z \|^2}{\| z \|^2} = 1.$$

Thus, $\frac{\overline{z}}{\|z\|^2}$ is the **multiplicative inverse** of z.

(M5) Finally, assume $z = a + bi, w_1 = c_1 + d_1 i, w_2 = c_2 + d_2 i$ are complex numbers with $a, b, c_1, d_1, c_2, d_2 \in \mathbb{R}$. We need to show that

$$(a + bi)[(c_1 + d_1 i) + (c_2 + d_2 i)] = (a + bi)(c_1 + d_1 i) + (a + bi)(c_2 + d_2 i). \quad (5.35)$$

The left hand side of the expression in (5.35) is

$$(a + bi)[(c_1 + c_2) + (d_1 + d_2)i] =$$

$$[a(c_1 + c_2) - b(d_1 + d_2)] + [a(d_1 + d_2) + b(c_1 + c_2)]i =$$

$$(ac_1 + ac_2 - bd_1 - bd_2) + (ad_1 + ad_2 + bc_1 + bc_2)i \quad (5.36)$$

The right hand side of the expression in (5.35) is

$$[(ac_1 - bd_1) + (ad_1 + bc_1)i] + [(ac_2 - bd_2) + (ad_2 + bc_2)i] =$$

$$(ac_1 - bd_1 + ac_2 - bd_2) + (ad_1 + bc_1 + ad_2 + bc_2)i \quad (5.37)$$

Since

$$ac_1 + ac_2 - bd_1 - bd_2 = ac_1 - bd_1 + ac_2 - bd_2$$

$$ad_1 + ad_2 + bc_1 + bc_2 = ad_1 + ad_2 + bc_1 + bc_2$$

it follows that $z(w_1 + w_2) = zw_1 + zw_2$. $\square$

The **complex field** is especially interesting and important because it is ***algebraically closed***. This means that every non-constant polynomial $f(z)$ with complex coefficients can be factored completely into linear factors. This is equivalent to the statement that every non-constant polynomial $f(z)$ with complex coefficients has a complex root. In particular, this is true of real polynomials.

Example 5.6.2. Determine the roots of the quadratic polynomial $x^2 + 6x + 11$.

We can use the **quadratic formula** which states that the roots of the quadratic polynomial $ax^2 + bx + c$ are $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$.

Applying the quadratic formula to $x^2 + 6x + 11$ we obtain the roots $\frac{-6 \pm \sqrt{36-44}}{2} = \frac{-6 \pm 2\sqrt{-2}}{2} = -3 \pm \sqrt{-2}$.

The square root of negative 2, $\sqrt{-2}$, can be expressed as a purely imaginary number: $\pm\sqrt{-2} = \pm\sqrt{2}\sqrt{-1} = \pm\sqrt{2}i$ since $i^2 = -1$ in the complex numbers. Therefore the roots of the polynomial $x^2 + 6x + 11$ are

$$-3 + \sqrt{2}i, -3 - \sqrt{2}i.$$

Notice that the roots form a pair of **complex conjugates**. This is always true of real quadratics which do not have real roots - the roots are a conjugate pair of complex numbers as can be seen from the quadratic formula.

If one looks at the definition of a **vector space** there is nothing that requires the scalars be real numbers, in fact, any field $\mathbb{F}$ can serve as the collection of scalars with the very same axioms. For completeness (and because it is so complicated) we write out the full definition here for further reference.

Definition 5.25. Assume $\mathbb{F}$ is a **field** . Let V be a set equipped with two operations: ***addition*** and ***scalar multiplication***. Addition is a rule (function) that associates to any pair $(\boldsymbol{u}, \boldsymbol{v})$ of elements from V a third element of V called the ***sum*** of $\boldsymbol{u}$ and $\boldsymbol{v}$ and denoted by $\boldsymbol{u} + \boldsymbol{v}$.

Scalar multiplication is a rule (function) which associates to each pair $(c, \boldsymbol{u})$ consisting of an element c of $\mathbb{F}$ (a scalar) and element $\boldsymbol{u}$ of V an element in V called the ***scalar multiple*** or ***scalar product*** of $\boldsymbol{u}$ by c and denoted by $c\boldsymbol{u}$. V is said to be a ***vector space over*** $\mathbb{F}$ if these operations satisfy the following axioms:

(A1) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ for every $\boldsymbol{u}, \boldsymbol{v} \in V$. **Addition is commutative**.

(A2) $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$ for every $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in V. **Addition is associative**.

(A3) There is a special element $\mathbf{0}$ called the **zero vector** such that $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$ for every $\boldsymbol{u} \in V$. This is the **existence of an additive identity**.

(A4) For every element $\boldsymbol{u}$ in V there is an opposite or negative element, denoted by $-\boldsymbol{u}$ such that $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$. This is the **existence of additive inverses**.

(M1) $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$ for every scalar a and vectors $\boldsymbol{u}, \boldsymbol{v} \in V$.

(M2) $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$ for all scalar a and b and vector $\boldsymbol{u}$.

(M3) $(ab)\boldsymbol{u} = a(b\boldsymbol{u})$ for all scalars a and b and vectors $\boldsymbol{u}$.

(M4) $1\boldsymbol{u} = \boldsymbol{u}$.

Definition 5.26. If V is a vector space over the **complex field**, $\mathbb{C}$, then we will say that V is a ***complex vector space***.

All the defiinitions that we introduced for real **vector space** carry over without modification to a **vector space** over a field $\mathbb{F}$: a sequence of vectors is **linearly independent**, the **span** of a sequence of vectors, **basis** of a vector space, **dimension** (of a finitely generated space) and so on. In particular, these can be formulated for **complex vector spaces**.

Example 5.6.3. 1) The complex analogue of $\mathbb{R}^n$ is $\mathbb{C}^n$ which consists of all columns of length n with complex entries. The addition is component-wise as is the scalar multiplication. This space has dimension n over $\mathbb{C}$, the **standard basis** is a **basis** .

2) $M_{m\times n}(\mathbb{C})$, the collection of all $m \times n$ matrices with the entries from the **complex field** is a **complex vector space** of **dimension** mn over $\mathbb{C}$.

Not only do all the fundamental definitions carry over but all the theorems we have proved do as well. For example, there is nothing special about the role of the reals in **Half is Good Enough Theorem** (5.3.8) or any of our other results. One can substitute any field for the reals and the theorems all hold with the same proof. The same is true of the many procedures we have developed. We illustrate with several examples.

Example 5.6.4. Determine if the sequence of vectors $\left(\begin{pmatrix} 1+i \\ 2 \\ 1-i \end{pmatrix}, \begin{pmatrix} -1+i \\ 2i \\ 1+i \end{pmatrix}\right)$ is **linearly independent** in $\mathbb{C}^3$.

Exercises like this in $\mathbb{R}^3$ involving two vectors were easy because one could eyeball and tell if the two vectors are multiples of each other; this is not as simple to recognize over the **complex field**. To determine if the sequence is **linearly independent** we make these the columns of a (complex) 3×2 matrix and apply **Gaussian elimination** and determine if each column is a **pivot column**.

The required matrix is $\begin{pmatrix} 1+i & -1+i \\ 2 & 2i \\ 1-i & 1+i \end{pmatrix}$

We begin by multiplying the first row by the multiplicative inverse of $1+i$ which is $\frac{1}{2}(1-i)$ in order to obtain a 1 in the (1,1)-position. The matrix we obtain is

$$\begin{pmatrix} 1 & i \\ 2 & 2i \\ 1-i & 1+i \end{pmatrix}$$

Next we add (-2) times the first row to the second row ($R_2 \to (-2)R_1 + R_2$) and add $-(1-i)$ times the first row to the third row ($R_3 \to [-(1-i)R_1 + R_3]$. The matrix we obtain is

$$\begin{pmatrix} 1 & i \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Therefore (surprisingly?) the two vectors are multiples of each other and the sequence is **linearly dependent**.

Example 5.6.5. Determine if the vector $\begin{pmatrix} 2i \\ 2+i \\ 3 \end{pmatrix}$ is in $Span\left(\begin{pmatrix} 1+i \\ 2 \\ 1-i \end{pmatrix}, \begin{pmatrix} 2i \\ 1+2i \\ 2-i \end{pmatrix}\right)$.

If so, then express $\begin{pmatrix} 2i \\ 2+i \\ 3 \end{pmatrix}$ as a **linear combination** of $\begin{pmatrix} 1+i \\ 2 \\ 1-i \end{pmatrix}$ and $\begin{pmatrix} 2i \\ 1+2i \\ 2-i \end{pmatrix}$.

We make an matrix with these three vectors and apply **Gaussian elimination** to obtain an **echelon form**. We then determine if the last column consisting of $\begin{pmatrix} 2i \\ 2+i \\ 3 \end{pmatrix}$ is a **pivot column**. If not then it is in the **span**, otherwise it is not. The matrix we require is

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

We begin by multiplying the first row by $(1+i)^{-1} = \frac{1}{2}(1-i)$ to obtain a one in the (1,1)-position. $(1-i)^{-1} = \frac{1+i}{2}$ and when we multiply the first row by this we get the matrix

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

We next add (-2) times the first row to the second row ($R_2 \rightarrow (-2)R_1 + R_2$) and $[-(1-i)]$ times the first row to the third row ($R_3 \rightarrow [-(1-i)]R_1 + R_3$). The resulting matrix is

$$\begin{pmatrix} 1 & 1+i & | & 1+i \\ 0 & -1 & | & -i \\ 0 & -i & | & 1 \end{pmatrix}$$

Next rescale the second row by multiplying by -1:

$$\begin{pmatrix} 1 & 1+i & | & 1+i \\ 0 & 1 & | & i \\ 0 & -i & | & 1 \end{pmatrix}$$

Add i times the second row to the third row to get complete the forward pass:

$$\begin{pmatrix} 1 & 1+i & | & 1+i \\ 0 & 1 & | & i \\ 0 & 0 & | & 0 \end{pmatrix}$$

At this point we see that the third column is not a **pivot column** and so $\begin{pmatrix} 2i \\ 2+i \\ 3 \end{pmatrix}$ is a **linear combination** of $\left(\begin{pmatrix} 1+i \\ 2 \\ 1-i \end{pmatrix}, \begin{pmatrix} 2i \\ 1+2i \\ 2-i \end{pmatrix} \right)$. We complete the **backward pass** in order to find a **linear combination**.

Add $[-(1+i)]$ times the second row to the first row ($R_1 \rightarrow R_1 + [-(1+i)]R_2$). This will give us a matrix in **reduced echelon form**.

$$\begin{pmatrix} 1 & 0 & | & 2 \\ 0 & 1 & | & i \\ 0 & 0 & | & 0 \end{pmatrix}$$

Therefore, $\begin{pmatrix} 2i \\ 2+i \\ 3 \end{pmatrix} = 2\begin{pmatrix} 1+i \\ 2 \\ 1-i \end{pmatrix} + i\begin{pmatrix} 2i \\ 1+2i \\ 2-i \end{pmatrix}$.

What You Can Now Do

1. **Multiply** two complex numbers.

2. Compute the **norm** of a complex number.

3. Compute the **multiplicative inverse** of a complex number.

4. Find the complex roots of a real quadratic polynomial.

5. Determine if a sequence of vectors in $\mathbb{C}^n$ is **linearly independent**. If the sequence is **linearly dependent** find a **non-trivial dependence relation**.

6. For a sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ from $\mathbb{C}^n$ and a complex n-vector $\boldsymbol{u}$, determine if $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ and, if so, then express $\boldsymbol{u}$ a (complex) **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$.

7. For a complex $m \times n$ matrix A find the **null space** of A.

Method (How To Do It)

Method 5.6.1. **Multiply** two complex numbers.

If the two complex numbers are $a + bi$ and $c + di$ with a, b, c, d real numbers then the product is $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$.

Example 5.6.6. Compute the following products:

a) $(2 + 3i)(4 - 5i)$

b) $(-1 + 4i)(2 + i)$

c) $(2 - 3i)^2$

d) $(1+i)^4$.

a) $(2+3i)(4-5i) = [(2)(4)-(3)(-5)] + [(2)(-5)+(3)(4)]i =$

$$[8+15] + [-10+12]i = 23+2i.$$

b) $(-1+4i)(2+i) = [(-1)(2)-(4)(1)] + [(-1)(1)+(4)(2)]i =$

$$[-2-4] + [-1+8]i = -6+7i.$$

c) $(2-3i)^2 = (2-3i)(2-3i) = [(2)(2)-(-3)(-3)] + [(2)(-3)+(2)(-3)]i =$

$$[4-9] + [-6-6]i = -5-12i.$$

d) $(1+i)^2 = (1+i)(1+i) = [(1)(1)-(1)(1)] + [(1)(1)+(1)(1)]i =$

$$[1-1] + [1+1]i = 2i.$$

$(1+i)^4 = (2i)^2 = 4i^2 = -4.$

Method 5.6.2. Compute the **norm** of a complex number.

If the complex numbers is $z = a+bi$ with a and b real numbers then $\| z \| = \sqrt{a^2+b^2}$.

Example 5.6.7. Find the **norm** of the each of the following complex numbers

a) $1+i$
b) $2-3i$
c) $-3+4i$
d) $4i$.

a) $\| 1+i \| = \sqrt{1^2+1^2} = \sqrt{2}$.

b) $\| 2-3i \| = \sqrt{2^2+(-3)^2} = \sqrt{13}$.

c) $\| -3+4i \| = \sqrt{(-3)^2+4^2} = \sqrt{25} = 5$.

d) $\| 4i \| = \sqrt{0^2+4^2} = \sqrt{16} = 4$.

Method 5.6.3. Compute the **multiplicative inverse** of a complex number z.

If $z = a + bi \neq 0$ with a, b real numbers then $z^{-1} = \frac{1}{\|z\|^2}\overline{z} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$.

Example 5.6.8. Find the **multiplicative inverse** of the following complex numbers:

a) $1 + i$
b) $3 - 4i$
c) $3i$
d) $1 - 2i$.

a) $(1+i)^{-1} = \frac{1}{2}(1-i) = \frac{1}{2} - \frac{1}{2}i$.

b) $(3-4i)^{-1} = \frac{1}{25}(3+4i) = \frac{3}{25} + \frac{4}{25}i$.

c) $(3i)^{-1} = \frac{1}{9}(-3i) = -\frac{1}{3}i$.

d) $(1-2i)^{-1} = \frac{1}{5}(1+2i) = \frac{1}{5} + \frac{2}{5}i$.

Method 5.6.4. Find the complex roots of a real quadratic polynomial.

Use the **quadratic formula**: The roots of the polynomial $ax^2 + bx + c$ are

$$\frac{-b+\sqrt{b^2-4ac}}{2a} \text{ and } \frac{-b-\sqrt{b^2-4ac}}{2a}$$

Example 5.6.9. Find the roots of the following real quadratic polynomials:

a) $x^2 - 6x + 14$
b) $x^2 + 4x - 5$
c) $x^2 + 8$
d) $x^2 + 7x + 13$.

a) The roots are $\frac{6\pm\sqrt{36-56}}{2} = \frac{6\pm\sqrt{-20}}{2} = \frac{6\pm2\sqrt{5}i}{2} = 3 \pm \sqrt{5}i$.

b) This actually has real roots since it can be factored $x^2 + 4x - 5 = (x+5)(x-1)$ and so the roots are -5 and 1. The quadratic formula yields the same roots:

$$\frac{-4 \pm \sqrt{16+20}}{2} = \frac{-4 \pm \sqrt{36}}{2} = \frac{-4 \pm 6}{2}.$$

So the roots are $\frac{-4+6}{2} = \frac{2}{2}$ = 1 and $\frac{-4-6}{2} = \frac{-10}{2}$ = -5.

c) The roots of x^2+8 are $\frac{0\pm\sqrt{-32}}{2} = \frac{\pm 4\sqrt{-2}}{2} = \pm 2\sqrt{2}i$.

d) The roots of $x^2+7x+13$ are $\frac{-7\pm\sqrt{49-52}}{2} = \frac{-7\pm\sqrt{-3}}{2} = -\frac{7}{2} \pm \frac{\sqrt{3}}{2}i$.

Method 5.6.5. Determine if a sequence $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ of vectors in $\mathbb{C}^n$ is **linearly independent**. If **linearly dependent**, find a **non-trivial dependence relation**.

Make the matrix $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_k)$. Apply **Gaussian elimination** (now using complex numbers) to obtain an **echelon form**. If every column is a **pivot column** then the sequence is **linearly independent**. If some column is not a **pivot column** continue with **Gaussian elimination** to obtain the **reduced echelon form** A'. Write out the **homogeneous linear system** (with complex coefficients) which has **coefficient matrix** equal to A' and find the **solution set**. Choose values for the parameters to obtain a particular solution.

Example 5.6.10. Determine if the sequence of vectors

$$\left(\begin{pmatrix} i \\ 2-i \\ i \end{pmatrix}, \begin{pmatrix} -1+i \\ 3+2i \\ 2+3i \end{pmatrix}, \begin{pmatrix} -1+2i \\ 6+5i \\ 2i \end{pmatrix}\right)$$

in $\mathbb{C}^3$ is **linearly independent**. If **linearly dependent**, find a **non-trivial dependence relation**.

We make the matrix $\begin{pmatrix} i & -1+i & -1+2i \\ 2-i & 3+2i & 6+5i \\ i & 2+3i & 2i \end{pmatrix}$ and apply **Gaussian elimination** .

Scale the first row by multiplying by $-i$ in order to get a one in the (1,1)-position. This yields the matrix $\begin{pmatrix} 1 & 1+i & 2+i \\ 2-i & 3+2i & 6+5i \\ i & 2+3i & 2i \end{pmatrix}$

Next add $[-(2-i)]$ times the first row to the second row and $(-i)$ times the first row to the third row. We obtain the matrix

$$\begin{pmatrix} 1 & 1+i & 2+i \\ 0 & i & 1+5i \\ 0 & 3+2i & 1 \end{pmatrix}$$

Scale the second row by multiplying by $-i$ to get a one in the (2,2)-position. We get the matrix

$$\begin{pmatrix} 1 & 1+i & 2+i \\ 0 & 1 & 5-i \\ 0 & 3+2i & 1 \end{pmatrix}$$

Add $[-(3+2i)]$ times the second row to the third row.

$$\begin{pmatrix} 1 & 1+i & 2+i \\ 0 & 1 & 5-i \\ 0 & 0 & -16-7i \end{pmatrix}$$

Every column is a **pivot column** and consequently the sequence of vectors is **linearly independent**.

Example 5.6.11. Determine if the sequence of vectors

$$\left(\begin{pmatrix} i \\ 2-i \\ i \end{pmatrix}, \begin{pmatrix} -1+i \\ 3+2i \\ 2+3i \end{pmatrix}, \begin{pmatrix} 2i \\ 5-2i \\ 2-i \end{pmatrix}\right)$$

in $\mathbb{C}^3$ is **linearly independent**. If **linearly dependent**, find a **non-trivial dependence relation**.

We make the matrix $\begin{pmatrix} i & -1+i & 2i \\ 2-i & 3+2i & 5-2i \\ i & 2+3i & 2-i \end{pmatrix}$ and apply **Gaussian elimination** .

Scale the first row by multiplying by $-i$ in order to get a one in the (1,1)-position. This yields the matrix $\begin{pmatrix} 1 & 1+i & 2 \\ 2-i & 3+2i & 5-2i \\ i & 2+3i & 2-i \end{pmatrix}$

Now we use elimination **elementary row operations** , adding $[-(2-i)]$ times the first row to the second row and $(-i)$ times the first row to the third row. We obtain the matrix

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

Scale the second row by $-i$ in order to get a one in the (2,2)-position. We get the matrix

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

Now add $[-(3+2i)]$ times the second row to the third row. The resulting matrix is

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

The last column is not a **pivot column** and therefore the sequence is **linearly dependent**. We complete the **Gaussian elimination** in order to get a **reduced echelon form** and then a **non-trivial dependence relation**.

Add $[-(1+i)]$ times the second row to the first row to get the following matrix in **reduced echelon form**:

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

The homogeneous linear system corresponding to this matrix is

$$\begin{array}{rcrcl} x & & + & (1+i)z & = & 0 \\ & y & - & iz & = 0 \end{array}$$

x, y are **leading variables** and z is a **free variable**. We set $z = t$, a complex parameter, and then we have $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -(1+i)t \\ it \\ 1 \end{pmatrix} = t \begin{pmatrix} -1-i \\ i \\ 1 \end{pmatrix}$.

Setting $t = 1$ we get the **non-trivial dependence relation**

$$-(1+i)\begin{pmatrix} i \\ 2-i \\ i \end{pmatrix} + i\begin{pmatrix} -1+i \\ 3+2i \\ 2+3i \end{pmatrix} + \begin{pmatrix} 2i \\ 5-2i \\ 2-i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Method 5.6.6. For a sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ from $\mathbb{C}^n$ and a vector $\boldsymbol{u}$ in $\mathbb{C}^n$ determine if $\boldsymbol{u} \in Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ and, if so, express $\boldsymbol{u}$ as a (complex) **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$.

Make the matrix $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_k |\ \boldsymbol{u})$. Use **Gaussian elimination** to obtain an **echelon form**. If the last column (the augmented column) is a **pivot column** then $\boldsymbol{u}$ is not in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$. If it is not a **pivot column** then it is in the span. In the latter case, continue with the **Gaussian elimination** to obtain the **reduced echelon form** A'. Write out the **inhomogeneous linear system** which has the matrix A' as its **augmented matrix**. Identify the **leading and free variables**. Set the free variables equal to (complex) parameters and express each of the variables in terms of the parameters to determine the **general solution** to the **linear system**. Any particular solution gives the coefficients for $\boldsymbol{u}$ as a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$.

Example 5.6.12. Determine if the vector $\begin{pmatrix} 1+3i \\ i \\ 2+i \end{pmatrix}$ is in

$Span\left(\begin{pmatrix} 1+i \\ 1-i \\ i \end{pmatrix}, \begin{pmatrix} 3+i \\ -2-i \\ 4+i \end{pmatrix}\right)$.

We make the matrix $\begin{pmatrix} 1+i & 3+i & | & 1+3i \\ 1-i & -2-i & | & i \\ i & 4+i & | & 2+i \end{pmatrix}$ and apply **Gaussian elimination** .
We begin by scaling the first row by $(1+i)^{-1} = \frac{1}{2}(1-i)$. We obtain the matrix

$$\begin{pmatrix} 1 & 2-i & | & 2+i \\ 1-i & -2+i & | & i \\ i & 4+i & | & 2+i \end{pmatrix}$$

We next add $[-(1-i)]$ times the first row to the second row and $(-i)$ times the first row to the third row to get the matrix

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

Scale the second row by $(-3+2i)^{-1}$ to get

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

Add $[-(3-i)]$ times the second row to the third row:

$$\begin{pmatrix} 1 & 2-i & | & 2+i \\ 0 & 1 & | & 1 \\ 0 & 0 & | & 0 \end{pmatrix}$$

This matrix is in **echelon form** and the last column is not a **pivot column** and therefore $\begin{pmatrix} 1+3i \\ i \\ 2+i \end{pmatrix}$ is in $Span\left(\begin{pmatrix} 1+i \\ 1-i \\ i \end{pmatrix}, \begin{pmatrix} 3+i \\ -2-i \\ 4+i \end{pmatrix}\right)$. We continue with the **Gaussian elimination** to get the **reduced echelon form**. Add $[-(2-i)]$ times the second row to the first row:

$$\begin{pmatrix} 1 & 0 & | & 2i \\ 0 & 1 & | & 1 \\ 0 & 0 & | & 0 \end{pmatrix}$$

There is a unique solution. We conclude that

$$\begin{pmatrix} 1+3i \\ i \\ 2+i \end{pmatrix} = (2i)\begin{pmatrix} 1+i \\ 1-i \\ i \end{pmatrix} + (1)\begin{pmatrix} 3+i \\ -2-i \\ 4+i \end{pmatrix}.$$

Method 5.6.7. For a complex $m \times n$ matrix A find the **null space** of A.

Apply **Gaussian elimination** to obtain the **reduced echelon form** R of A. If every column is a **pivot column** then the **null space** of A consists of only the **zero vector**, $\mathbf{0}_n$. Otherwise, write out the **homogeneous linear system** which has the matrix R as its **coefficient matrix**. Identity the **leading and free variables**. Assign (complex) parameters to the **free variables** and express each variable in terms of these parameters. The general solution can be expressed as a **linear combination** of complex vectors where the coefficients are the parameters. These vectors form a **basis** for the **null space** of R and A.

Example 5.6.13. Find the **null space** of the complex matrix $A = \begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix}$

We scale the first row to obtain the matrix $\begin{pmatrix} 1 & i \\ -1 & -i \end{pmatrix}$ and then add the first row to the second row to obtain the **reduced echelon form** $\begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix}$. The corresponding **homogeneous linear system** (equation in this case) is

$$x \ + \ iy \ = \ 0$$

The variable x is the leading variable and y is the free. We set $y = t$, a complex parameter and express x and y in terms of t. we get $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -ti \\ t \end{pmatrix} = t\begin{pmatrix} -i \\ 1 \end{pmatrix}$.

Therefore, the **null space** of A is $Span\left(\begin{pmatrix} -i \\ 1 \end{pmatrix}\right)$.

Example 5.6.14. Find the **null space** of the complex matrix $A = \begin{pmatrix} 1-i & 2 \\ -1 & -1-i \end{pmatrix}$.

We begin by scaling the first row so the (1,1)-entry becomes a one. Thus, we multiply by first row by $\frac{1}{2}(1+i)$. The matrix we obtain is $\begin{pmatrix} 1 & 1+i \\ -1 & -1-i \end{pmatrix}$. We then add the first row to the second row to obtain the **reduced echelon form** , $R = \begin{pmatrix} 1 & 1+i \\ 0 & 0 \end{pmatrix}$. The **homogeneous linear system (equation)** corresponding to R is

$$x \quad + \quad (1+i)y \quad = 0 \quad .$$

The variable y is free. We set it equal to t, a complex paramete,r and express x and y in terms of t. Thus,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -(1+i)t \\ t \end{pmatrix} = t\begin{pmatrix} -1-i \\ 1 \end{pmatrix}$$

and therefore $null(A) = Span\left(\begin{pmatrix} -1-i \\ 1 \end{pmatrix}\right)$.

Example 5.6.15. Find the **null space** of the complex matrix

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

We apply **Gaussian elimination** . First, scale the first row by the factor $(-1+i)^{-1} = -\frac{1+i}{2}$ to get the matrix

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

Next add (-2) times the first row to the third row to obtain the matix

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

Now scale the second row by $(-1+i)^{-1} = -\frac{1+i}{2}$. This produces the matrix

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

Add $(2-i)$ times the second row to the third row. We then get the matrix

$$\begin{pmatrix} 1 & -\frac{1+i}{2} & 0 \\ 0 & 1 & -1-i \\ 0 & 0 & 0 \end{pmatrix}$$

Finally, add $\frac{1+i}{2}$ times the second row to the first to get the **reduced echelon form** , R of the matrix A :

$$R = \begin{pmatrix} 1 & 0 & -i \\ 0 & 1 & -1-i \\ 0 & 0 & 0 \end{pmatrix}$$

The homogeneous linear system with this as its coefficient matrix is

$$\begin{array}{ccccc} x & & - & iz & = & 0 \\ & y & - & (1+i)z & = & 0 \end{array}.$$

The variables x and y are **leading variables** and z is a **free variable**. We set $z = t$, a complex parameter, and solve for all the variables in terms of t :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} it \\ (1+i)t \\ t \end{pmatrix} = t \begin{pmatrix} i \\ 1+i \\ 1 \end{pmatrix}.$$

Therefore, $null(A) = Span\left(\begin{pmatrix} i \\ 1+i \\ 1 \end{pmatrix}\right)$.

Exercises

In exercises 1 - 4 compute the given product of complex numbers. See **Method** (5.6.1).

1. $(2+3i)i$.

2. $(-2+i)(3+4i)$

3. $(-3-5i)(3-5i)$

4. $(2+5i)(1-2i)$

In exercises 5 - 9 find the norm of the given complex number. See **Method** (5.6.2).

5. $1-2i$

6. $1-3i$

7. $-2+5i$

8. $12-5i$

9. $6i$

In exercises 10 - 13 find the inverse of the given complex number. See **Method** (5.6.3).

10. $1+3i$

11. $-1-i$

12. $-3i$

13. $2-3i$

In exercises 14 - 18 find the roots of the given real quadratic polynomial. See **Method** (5.6.4).

14. $x^2-8x+25$

15. x^2+5

16. x^2+4x+2

17. x^2-3x+3

18. $x^2-6x+11$

In exercises 19 - 24 determine if the given sequence of vectors from $\mathbb{C}^3$ is **linearly independent**. If **linearly dependent**, find a **non-trivial dependence relation**. See **Method** (5.6.5).

19. $\left(\begin{pmatrix} 1 \\ 1+i \\ 2+i \end{pmatrix}, \begin{pmatrix} 1+i \\ 2i \\ 1+3i \end{pmatrix}\right)$

20. $\left(\begin{pmatrix}1\\i\\0\end{pmatrix}, \begin{pmatrix}i\\1\\1+i\end{pmatrix}, \begin{pmatrix}0\\-1+i\\-1+2i\end{pmatrix}\right)$

21. $\left(\begin{pmatrix}1\\-1+i\\1-i\end{pmatrix}, \begin{pmatrix}2\\-3+i\\3\end{pmatrix}, \begin{pmatrix}0\\2\\-3-i\end{pmatrix}\right)$

22. $\left(\begin{pmatrix}1\\i\\0\end{pmatrix}, \begin{pmatrix}0\\1\\-i\end{pmatrix}, \begin{pmatrix}2\\1+2i\\1-i\end{pmatrix}, \begin{pmatrix}2+i\\1+2i\\1-2i\end{pmatrix}\right)$

23. $\left(\begin{pmatrix}1\\i\\1+i\end{pmatrix}, \begin{pmatrix}0\\1\\i\end{pmatrix}, \begin{pmatrix}1\\i\\1+i\end{pmatrix}\right)$

24. $\left(\begin{pmatrix}1+i\\1\\0\end{pmatrix}, \begin{pmatrix}0\\i\\1+i\end{pmatrix}, \begin{pmatrix}1\\0\\i\end{pmatrix}\right)$

In exercises 25 - 28 determine if the complex vector $\boldsymbol{u}$ is a **linear combination** of the complex sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2)$. See **Method** (5.6.6).

25. $\boldsymbol{v}_1 = \begin{pmatrix}i\\1-i\\-1\end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix}2+3i\\-3-2i\\1-i\end{pmatrix}, \boldsymbol{u} = \begin{pmatrix}1\\-1+i\\-i\end{pmatrix}$

26. $\boldsymbol{v}_1 = \begin{pmatrix}1\\1+i\\i\end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix}1\\2+i\\1+i\end{pmatrix}, \boldsymbol{u} = \begin{pmatrix}1\\1\\1\end{pmatrix}$

27. $\boldsymbol{v}_1 = \begin{pmatrix}1+i\\1\\0\end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix}1\\i\\1+2i\end{pmatrix}, \boldsymbol{u} = \begin{pmatrix}i\\1\\-i\end{pmatrix}$

28. $\boldsymbol{v}_1 = \begin{pmatrix}i\\1+i\\i\end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix}1+i\\2+3i\\-2+i\end{pmatrix}, \boldsymbol{u} = \begin{pmatrix}i\\1+4i\\-3+i\end{pmatrix}$

In exercises 29 - 34 find the **null space** of the given complex matrix.

29. $\begin{pmatrix}-i & 1\\-1 & -i\end{pmatrix}$

30. $\begin{pmatrix}1+i & 2\\i & 1+i\end{pmatrix}$

31. $\begin{pmatrix}-1+i & 2\\-1 & 1+i\end{pmatrix}$

32. $\begin{pmatrix} 1 & i & 1+i \\ i & -1+i & 2 \end{pmatrix}$

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

34. $\begin{pmatrix} 1 & i & 0 \\ 2-i & -i & 2 \\ -2 & 0 & -1-i \end{pmatrix}$

Challenge Exercises (Problems)

1. For a complex vector $\boldsymbol{v}$, let $\overline{\boldsymbol{v}}$ be the result of taking the complex conjugate of all the components of $\boldsymbol{v}$. For $\boldsymbol{v} \in \mathbb{C}^n$ set $\boldsymbol{v}_{\mathbb{R}} = \frac{\boldsymbol{v}+\overline{\boldsymbol{v}}}{2}$ and $\boldsymbol{v}_I = \frac{\boldsymbol{v}-\overline{\boldsymbol{v}}}{2i}$. Prove that $\boldsymbol{v}_{\mathbb{R}}$ and $\boldsymbol{v}_I$ are in $\mathbb{R}^n$ and that $\boldsymbol{v}_{\mathbb{R}} + i\boldsymbol{v}_I = \boldsymbol{v}$.

2. Let A be a real $m \times n$ matrix and let $\boldsymbol{v} \in \mathbb{C}^n$. Prove that $\boldsymbol{v}$ is in the **null space** of A if and only if $\boldsymbol{v}_{\mathbb{R}} \in null(A)$ and $\boldsymbol{v}_I \in null(A)$.

3. For a complex matrix A let $\overline{A}$ be the matrix obtained by replacing each entry of A by its complex conjugate.

a) Let $\boldsymbol{z}, \boldsymbol{w}$ be vectors in $\mathbb{C}^n$. Prove that $\overline{\boldsymbol{z}^{Tr}\boldsymbol{w}} = \overline{\boldsymbol{z}}^{Tr}\overline{\boldsymbol{w}}$.

b) Let A be a complex $m \times n$ matrix and B be a complex $n \times p$ matrix. Prove that $\overline{AB} = \overline{A}\,\overline{B}$.

4. Prove that $\boldsymbol{v} \in null(A)$ if and only if $\overline{\boldsymbol{v}} \in null(\overline{A})$.

5. Prove parts 1) and 2) of **Theorem** (5.6.2).

Quiz Solutiions

1. $-2+5i$

Not right, see **definition of addition of complex numbers**.

2. $-11-2i$

Not right, see **definition of addition of complex numbers**.

3. -1

Not right, see **definition of addition of complex numbers**.

4. $-6+2i$

Not right, see **definition of addition of complex numbers**.

5. 10

Not right, see **definition of multiplication of complex numbers**.

6. $2i$

Not right, see **definition of multiplication of complex numbers**.

7. $14 + 5i$

Not right, see **definition of multiplication of complex numbers**.

5.7. Vector Spaces Over Finite Fields

In this section we introduce the notion of a finite **field** and work through some examples. We then discuss **vector spaces** over finite fields. In the next section of this chapter vector spaces over finite fields is applied to the construction of error correcting codes, which are exceedingly important to modern communications.

Am I Ready for This Material

Essential to understanding the material of this section are a mastery of the following previously introduced concepts:

a **field**

a **vector space** over a field $\mathbb{F}$

linear combination of a sequence of vectors in a vector space

subspace of a vector space

span of a sequence or a set of vectors from a vector space

spanning sequence of a subspace of a vector space

linearly dependent sequence of vectors from a vector space

linearly independent sequence of vectors from a vector space

basis of a vector space

finitely generated vector space

dimension of a finitely generated vector space

finite dimensional vector space

Quiz

1. Solve the linear equation $3x + 7 = 7x - 13$.

2. Find the remainder of 16×13 when divided by 17.

3. Find the **null space** of the matrix $A = \begin{pmatrix} 1 & 3 & 1 & 2 & 2 \\ 1 & 2 & 2 & 1 & 3 \\ 1 & 2 & 2 & 1 & 2 \\ 2 & 4 & 4 & 2 & 6 \\ 1 & 3 & 1 & 2 & 1 \end{pmatrix}$

4. Find the **solution set** to the **linear system**

$$\begin{array}{rcrcrcrcr} x_1 & - & 2x_2 & - & x_3 & + & 2x_4 & = & 1 \\ 2x_1 & - & 3x_2 & - & 3x_3 & + & 4x_4 & = & 2 \\ x_1 & - & 3x_2 & & & + & 2x_4 & = & 1 \end{array}$$

Quiz Solutions

New Concepts

We introduce one new concept in this section, namely, a **finite field**. Such an object is a **field** which has only a finite number of elements (unlike the **rational and real fields** and the **complex field**). It is possible to construct such a **field** with exactly p elements where p is any prime number, in fact, finite fields exist with p^k elements where p is any prime and k is a natural number. Demonstrating the existence of such fields is beyond the scope of this material and will have to await results from **ring theory** which are usually proved in a course in **abstract algebra**.

Theory (Why It Works)

We have seen in the previous section that the collection of scalars in a **vector space** need not be the **rational or real fields** but can be an artbitrary **field**. The **complex field** is an example and complex **vector spaces** have important applications in science and techology.

A **field** need not be infinite. In fact, **finite fields** exist in abundance and **vector spaces** over finite fields also have many applications - not only in mathematics but also in technology. For example, **vector spaces** over finite fields are applied to the construction of **error correcting codes** which are essential for reliably sending information digitally over "noisy" channels. We will discuss such an application in the next section. In the present section we will define a finite field, work out several examples and then demonstrate that the algorithms and methods we originally developed for real vector spaces apply to **vector spaces** over finite fields as well.

Definition 5.27. A ***finite field*** is a field $(\mathbb{F}, +, \cdot, 0, 1)$ in which the set $\mathbb{F}$ is finite.

We now describe a collection of finite fields $\mathbb{F}_p$ where p is a prime number.

Definition 5.28. Let p be a **prime number**. The set unlying $\mathbb{F}_p$ will consist of $\{0, 1, \ldots, p-1\}$. For $a, b \in \mathbb{F}_p$ the sum $a \oplus_p b$ is defined as follows: if $a + b < p$ then $a \oplus_p b = a + b$. If $a + b \geq p$ then $a \oplus_p b = (a + b - p)$.

The product of a and b, denoted by $a \otimes_p b$ is defined similarly: If $ab < p$ then $a \otimes_p b = ab$. If $ab \geq p$ then find the **remainder** r of ab when it is divided by p. Then $a \otimes_p b = r$. The set $\{0, 1, \ldots, p-1\}$ with this addition and multiplication together with 0 as the additive identity and 1 as the multiplicative identity is the finite field $\mathbb{F}_p$.

In the definition we refer to $\mathbb{F}_p$ as a **field** and this requires proof. However, the proof requires many results found in a first or second course in **abstract algebra**. and it is beyond what we can do here. Therefore the reader will have to be satisfied with the assurance that this is, indeed, a **field**. However, we will give some small examples which can be shown to be fields through a sufficient number of calculations. We will specify the examples by writing down the addition and multiplication tables for the field. These are tables which have rows and columns indexed by the elements of the field, namely the symbols $0, 1, \ldots, p-1$. In the addition table, at the intersection of the row indexed by the element a and the column indexed by the element b we place the element $a \oplus_p b$. The multiplication table is constructed in a similar way.

Example 5.7.1. When $p = 2$ the field consists of just two elements, 0 and 1. The addition and multiplication are shown in the following tables.

$\oplus_2$	0	1
0	0	1
1	1	0

$\otimes_2$	0	1
0	0	0
1	0	1

Remark 5.14. The product of any element with the additive identity, zero, is always zero and consequently, hereafter we will not include zero in the multiplication table for a field.

Example 5.7.2. When $p = 3$ the field consists of the elements 0,1,2. The addition and multiplication tables are as shown.

$\oplus_3$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\otimes_3$	1	2
1	1	2
2	2	1

Notice that 2 is its own **multiplicative inverse** since $2 \otimes_3 2 = 1$.

Example 5.7.3. For $p = 5$ the elements are 0,1,2,3,4. The addition and multiplication tables are as follows.

$\oplus_5$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\otimes_5$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Notice that 4 is its own **multiplicative inverse** since $4 \otimes_5 4 = 1$ and that 2 and 3 are inverses of each other.

Other possibilities exist in addition to the fields $\mathbb{F}_p$ for p a finite number, but they all have one thing in common: the number of elements in such a field is always a power of a prime number, and, in fact, for every prime p and positive integer n there is a field $\mathbb{F}_{p^n}$ with p^n elements. The construction of such a field, in general, requires many notions and ideas from the branch of mathematics known as **abstract algebra**. and cannot be given here. However, the field $\mathbb{F}_{p^n}$ is constructed from $\mathbb{F}_p$ in a way similar to the construction of the complex field, $\mathbb{C}$, as an extension of the real field, $\mathbb{R}$, that is, obtained by adjoining a root of some irreducible polynomial.

In the absence of a general construction we define $\mathbb{F}_4$ concretely by specifying its elements and writing down the addition and multiplication table.

Example 5.7.4. The elements of $\mathbb{F}_4$ will be denoted by $0, 1, \omega$ and $\omega + 1$. The table for addition and multiplication follow.

$\oplus_4$	0	1	ω	$\omega+1$
0	0	1	ω	$\omega+1$
1	1	0	$\omega+1$	ω
ω	ω	$\omega+1$	0	1
$\omega+1$	$\omega+1$	ω	1	0

$\otimes_4$	1	ω	$\omega+1$
1	1	ω	$\omega+1$
ω	ω	$\omega+1$	1
$\omega+1$	$\omega+1$	1	ω

Notice that for any $a \in \mathbb{F}_4$, the sum of a with a is zero: $a \oplus_4 a = 0$.

We also point out that $\omega \otimes_4 (\omega+1) = 1$ so that ω and $\omega+1$ are **multiplicative inverses** of each other. Finally, observe that $\omega^3 = (\omega + 1) \otimes_4 \omega = 1$ and likewise $(\omega + 1)^3 = \omega \otimes_4 (\omega + 1) = 1$ and therefore ω and $\omega + 1$ are cubic roots of 1.

As with any **field** , **vector spaces** can be defined over **finite fields**. As we will see in the next section, **vector spaces** over **finite fields** are the foundation for the construction of **error correction codes** which are essential to the operation of modern digital communication. Some examples of vector spaces over finite fields follow. Throughout these examples $\mathbb{F}_q$ will denote an arbitrary **finite field** with q elements.

Example 5.7.5. Let $\mathbb{F}_q^n$ consists of all $n \times 1$ columns with entries in $\mathbb{F}_q$. Addition and scalar multiplication are done component-wise:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 \oplus_q b_1 \\ a_2 \oplus_q b_2 \\ \vdots \\ a_n \oplus_q b_n \end{pmatrix}, c \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} c \otimes_q a_1 \\ c \otimes_q a_2 \\ \vdots \\ c \otimes_q a_n \end{pmatrix}$$

This is a **vector space** of **dimension** n.

Example 5.7.6. Let $M_{m\times n}(\mathbb{F}_q)$ consist of all $m \times n$ matrices with entries in $\mathbb{F}_q$. With component wise addition and scalar multiplication. This is a **vector space** of **dimension** mn.

Example 5.7.7. Let $(\mathbb{F}_q)_n[x]$ denote all polynomials with coefficients in $\mathbb{F}_q$ and of degree at most n in an indeterminate x. Thus, $(\mathbb{F}_q)_n[x]$ consists of all expressions of the form

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where $a_0, a_1, \ldots, a_n \in \mathbb{F}_q$. We add in the obvious way:

$$(a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n) =$$

$$(a_0 \oplus_q b_0) + (a_1 \oplus_q b_1)x + \cdots + (a_n \oplus b_n)x^n.$$

Similarly, for scalar multiplication:

$$c(a_0 + a_1x + \cdots + a_nx^n) = (c \otimes_q a_0) + (c \otimes_q a_1)x + \cdots + (c \otimes_q a_n)x^n$$

All the theorems of this chapter and the procedures/methods/algorithms developed throughout the book apply to a **vector space** over a **finite field**. Some of these are illustrated below in the method subsection.

What You Can Now Do

1) Given the addition and multiplication tables for a **finite field** $\mathbb{F}_q$, solve the **linear system** with **augmented matrix** $(A|\boldsymbol{b})$, where A is an $m \times n$ matrix with entries in $\mathbb{F}_q$ and $\boldsymbol{b}$ is an $m-$vector with entries in $\mathbb{F}_q$.

2) Determine if a vector $\boldsymbol{b}$ from $\mathbb{F}_q^m$ is in the **span** of the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ from $\mathbb{F}_q^m$. If $\boldsymbol{b}$ is in $Span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ express $\boldsymbol{b}$ as a **linear combination** of $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$.

3) Given a sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ in $\mathbb{F}_q^n$ determine the **dimension** of $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$.

Method (How To Do It)

Method 5.7.1. Given the addition and multiplication tables for a **finite field** $\mathbb{F}_q$, solve the **linear system** with **augmented matrix** $(A|\boldsymbol{b})$, where A is an $m \times n$ matrix with entries in $\mathbb{F}_q$ and $\boldsymbol{b}$ is an $m-$vector with entries in $\mathbb{F}_q$.

Use **Gaussian elimination** to obtain the **reduced echelon form** R of $(A|\boldsymbol{b})$. If the last (augmented) column of R is a **pivot column** then the system is **inconsistent**. Otherwise, write down the **linear system** with **augmented matrix** R. Determine the **leading and free variables**. Set the **free variables** equal to parameters and solve for all the variables in terms of these parameters.

Example 5.7.8. Find the **solution set** to the **linear system** with **augmented matrix**

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & | & 0 \\ 1 & 2 & 1 & 0 & 1 & | & 0 \\ 2 & 1 & 2 & 2 & 1 & | & 0 \\ 1 & 2 & 1 & 2 & 1 & | & 1 \end{pmatrix}$$

where the entries are in $\mathbb{F}_3$.

We will just use the notation + for $\oplus_3$ throughout. Recall that

$$1 + 1 = 2, 1 + 2 = 0, 1 \times 1 = 2 \times 2 = 1$$

We begin by adding appropriate multiplies of the first row to the rows below it to make the other entries in the first column zero. The operations are: $R_2 \to 2R_1 + R_2, R_3 \to$

$R_1 + R_3, R_4 \to 2R_1 + R_4$. After performing these **elementary row operations** we obtain the matrix:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & | & 0 \\ 0 & 1 & 1 & 2 & 1 & | & 0 \\ 0 & 2 & 2 & 0 & 1 & | & 0 \\ 0 & 1 & 1 & 1 & 1 & | & 1 \end{pmatrix}$$

We next perform the following elimination-type **elementary row operations** making use of the second row: $R_3 \to R_2 + R_3, R_4 \to 2R_2 + R_4$. The matrix we obtain is

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & | & 0 \\ 0 & 1 & 1 & 2 & 1 & | & 0 \\ 0 & 0 & 0 & 2 & 2 & | & 0 \\ 0 & 0 & 0 & 2 & 0 & | & 1 \end{pmatrix}$$

We scale the third row, multiplying by 2, $R_3 \to 2R_3$ which yields the matrix

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & | & 0 \\ 0 & 1 & 1 & 2 & 1 & | & 0 \\ 0 & 0 & 0 & 1 & 1 & | & 0 \\ 0 & 0 & 0 & 2 & 0 & | & 1 \end{pmatrix}$$

Add the third row to the fourth $(R_4 \to R_3 + R_4)$ and this will complete the forward pass:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & | & 0 \\ 0 & 1 & 1 & 2 & 1 & | & 0 \\ 0 & 0 & 0 & 1 & 1 & | & 0 \\ 0 & 0 & 0 & 0 & 1 & | & 1 \end{pmatrix}$$

We see at this point that the system is **consistent**. We complete the **backward pass** to get the **reduced echelon form**. We begin with the elimination **elementary row operations** $R_2 \to R_2 + 2R_4, R_3 \to R_3 + 2R_4$:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & | & 0 \\ 0 & 1 & 1 & 2 & 0 & | & 2 \\ 0 & 0 & 0 & 1 & 0 & | & 2 \\ 0 & 0 & 0 & 0 & 1 & | & 1 \end{pmatrix}$$

Next add twice the third row to the first, $R_1 \to R_1 + 2R_3$, and add the third row to the second row, $R_2 \to R_2 + R_3$ to get

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

Finally, add twice the second row to the first row, $R_1 \to R_1 + 2R_2$ to obtain the **reduced echelon form**:

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

The **linear system** with this as its **augmented matrix** is

$$\begin{array}{lllllll} x_1 & & + 2x_3 & & & = & 0 \\ & x_2 & + x_3 & & & = & 1 \\ & & & x_4 & & = & 2 \\ & & & & x_5 & = & 1 \end{array}$$

The are four **leading variables** (x_1, x_2, x_4, x_5) and one **free variable** (x_3). We set $x_3 = t$ and get the general solution:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} t \\ 1+2t \\ t \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \\ 1 \end{pmatrix} + t \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Method 5.7.2. Determine if a vector $\boldsymbol{b}$ from $\mathbb{F}_q^m$ is in the **span** of the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ from $\mathbb{F}_q^m$. If $\boldsymbol{b}$ is in $Span(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ express $\boldsymbol{b}$ as a **linear combination** of $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$.

Let A be the matrix with columns the sequence of vectors $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$. Use **Method** (5.7.1) to determine if the **linear system** with **augmented matrix** $(A|\boldsymbol{b})$ is **consistent**. If the **linear system** is **consistent** continue with **Method** (5.7.1) to find a **solution** to the **linear system**. This will be an a vector $\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ from $\mathbb{F}_q^n$ such that $\boldsymbol{b} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$.

Example 5.7.9. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix}$ be vectors in $\mathbb{F}_5^3$. Determine which, if any, of the standard basis vectors is in the **span** of $(\boldsymbol{v}_1, \boldsymbol{v}_2)$, $Span(\boldsymbol{v}_1, \boldsymbol{v}_2)$.

Let $A = \begin{pmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 1 \end{pmatrix}$. We need to determine which, if any, of the **linear systems** with **augmented matrix** $(A|\boldsymbol{e}_1), (A|\boldsymbol{e}_2)$, or $(A|\boldsymbol{e}_3)$ is **consistent**.

$(A|\boldsymbol{e}_1) = \begin{pmatrix} 1 & 1 & | & 1 \\ 2 & 4 & | & 0 \\ 3 & 1 & | & 0 \end{pmatrix}$ We begin by performing the following operations: $R_2 \to 3R_1 + R_2, R_3 \to 2R_1 + R_3$:

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

We scale the second row by 3 to get

$$\begin{pmatrix} 1 & 1 & | & 1 \\ 0 & 1 & | & 4 \\ 0 & 3 & | & 2 \end{pmatrix}$$

We then add 2 times the second row to the third row:

$$\begin{pmatrix} 1 & 1 & | & 1 \\ 0 & 1 & | & 4 \\ 0 & 0 & | & 0 \end{pmatrix}$$

This completes the forward pass and we see the **linear system** with this matrix is **consistent** since the last column is not a **pivot column**. So $\boldsymbol{e}_1$ is in the **span** of $(\boldsymbol{v}_1, \boldsymbol{v}_2)$. We can get the **reduced echelon form** in one operation and then express $\boldsymbol{e}_1$ as a **linear combination** of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Add 4 times row two to row one, $R_1 \to R_1 + 4R_2$:

$$\begin{pmatrix} 1 & 0 & | & 2 \\ 0 & 1 & | & 4 \\ 0 & 0 & | & 0 \end{pmatrix}$$

Thus, we see that $\boldsymbol{e}_1 = 2\boldsymbol{v}_1 + 4\boldsymbol{v}_2$. We leave the others as an exercise.

Method 5.7.3. Given a sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$ in $\mathbb{F}_q^n$ determine the **dimension** of $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$.

Make the matrix $(\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \ldots \ \boldsymbol{v}_n)$. Use **Gaussian elimination** to find an **echelon form**. The **dimension** of $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ is the number of **pivot columns** of this **echelon form**.

Example 5.7.10. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 4 \\ 1 \\ 4 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 4 \\ 4 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}, \boldsymbol{v}_5 = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 3 \end{pmatrix}$ be vectors in $\mathbb{F}_5^4$. Determine the **dimension** of $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5)$.

We need to compute an **echelon form** of the matrix $(\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \boldsymbol{v}_3 \ \boldsymbol{v}_4 \ \boldsymbol{v}_5) =$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 2 & 4 & 1 & 1 & 2 \\ 3 & 1 & 4 & 1 & 0 \\ 4 & 4 & 4 & 2 & 3 \end{pmatrix}$$

We perform the following **elementary row operations**: $R_2 \to 3R_1 + R_2, R_3 \to 2R_1 + R_3, R_4 \to R_1 + R_4$ to obtain the matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 2 & 4 & 4 & 2 \\ 0 & 3 & 1 & 3 & 0 \\ 0 & 0 & 0 & 3 & 3 \end{pmatrix}$$

Scale the second row, $R_3 \to 3R_2$ to get

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

Next, perform $R_3 \to 2R_2 + R_3$ to get the matrix

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

After performing $R_3 \to 3R_3$ and then $R_4 \to 2R_3 + R_4$ we get the following matrix **echelon form** :

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

This matrix has three **pivot columns** , namely, columns 1, 2 and 4. Therefore $dim(Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5)) = 3$. In fact, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_4)$ is a **basis** for the span of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5)$.

Exercises

In exercises 1 - 6 determine the **solution** to the given **linear equation** with entries in the given finite field, $\mathbb{F}_q$.

1. $2x + 1 = 2$ with $q = 3$.

2. $2x + 1 = 4$ with $q = 5$.

3. $4x + 3 = 2$ with $q = 5$.

4. $3x + 2 = 0$ with $q = 5$.

5. $\omega x + 1 = 0$ with $q = 4$.

6. $(\omega + 1)x = \omega$ with $q = 4$.

In exercises 7 - 10 determine the **solution set** of the **linear system** with the given **augmented matrix** with entries in the specified finite field. See **Method** (5.7.1).

7. $\begin{pmatrix} 1 & 0 & 1 & 0 & | & 0 \\ 1 & 1 & 0 & 0 & | & 1 \\ 1 & 1 & 1 & 1 & | & 0 \end{pmatrix}$ with entries in $\mathbb{F}_2$.

8. $\begin{pmatrix} 1 & 1 & 1 & 1 & | & 0 \\ 1 & 2 & 0 & 0 & | & 1 \\ 1 & 1 & 1 & 2 & | & 1 \\ 0 & 2 & 1 & 0 & | & 1 \end{pmatrix}$ with entries in $\mathbb{F}_3$.

9. $\begin{pmatrix} 1 & 1 & 1 & 1 & | & 1 \\ 1 & 2 & 3 & 4 & | & 0 \\ 1 & 4 & 4 & 1 & | & 0 \\ 1 & 3 & 2 & 4 & | & 0 \end{pmatrix}$ with entries in $\mathbb{F}_5$.

10. $\begin{pmatrix} 1 & 1 & 1 & | & 0 \\ 1 & \omega & \omega+1 & | & 1 \\ 1 & \omega+1 & \omega & | & 0 \end{pmatrix}$ with entries in $\mathbb{F}_4$

In exercises 11 - 13 determine the **rank** of the given matrix with entries in the specified field. See **Method** (5.7.2).

11. $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ with entries in $\mathbb{F}_2$.

12. $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 \\ 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 2 \end{pmatrix}$ with entries in $\mathbb{F}_3$.

13. $\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 4 & 1 \\ 1 & 3 & 2 & 2 & 4 \\ 1 & 4 & 1 & 0 & 4 \end{pmatrix}$ with entries in $\mathbb{F}_5$.

In exercises 14 - 17 answer true or false and give an explanation.

14. For any positive integer n there is a **finite field** with n elements.

15. If V is a **vector space** of **dimension** 4 over the field $\mathbb{F}_2$ then V has 16 vectors.

16. In the field $\mathbb{F}_4$ every element α satisfies $\alpha + \alpha = 0$.

17. In a **vector space** over $\mathbb{F}_3$ any **subspace** of dimension two has six elements.

Challenge Exercises (Problems)

For a vector $\boldsymbol{x} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ in $\mathbb{F}_q^n$ define the ***weight*** of $\boldsymbol{x}, wt(\boldsymbol{x})$ to be the number of i such that $a_i \neq 0$.

For example, the vector $\begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \\ 3 \\ 0 \\ 4 \end{pmatrix}$ in $\mathbb{F}_5^7$ has weight 4.

1. Let $A = (\boldsymbol{v}_1\ \boldsymbol{v}_2 \ldots\ \boldsymbol{v}_n)$ be a matrix with entries in $\mathbb{F}_q$ and assume no two of the columns $\boldsymbol{v}_j$ of A are multiples of each other. Let $\boldsymbol{x}$ be a nonzero vector in $null(A)$. Prove that $wt(\boldsymbol{x}) \geq 3$.

2. Let $A = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$ have entries in $\mathbb{F}_2$.

a) Prove the **rank** of A is 3 and that **dimension** of the **null space** of A is 4.

b) Prove that for any nonzero vector $\boldsymbol{x} \in null(A)$ that $wt(\boldsymbol{x}) \geq 3$.

3. Write out the addition and multiplication tables for the field $\mathbb{F}_7$ where addition and multiplication are "modulo" 7. For example $4 \oplus_7 5 = 2$ and $3 \otimes_7 4 = 5$.

Quiz Solutions

1. $x = 5$. Not right, see **Example** (1.1.1).

2. The remainder when 16×13 is divided by 17 is 4. Not right, see the **division algorithm**.

3. The **null space** of the matrix is $Span\left(\begin{pmatrix} -4 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}\right)$. Not right, see **Method** (3.2.2).

4. The **solution set** to the **linear system** is $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + Span\left(\begin{pmatrix} 3 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right)$. Not right, see **Method** (3.2.3).

5.8. How to Use it: Error Correcting Codes

This section demonstrates how a **vector space** over a **finite field** can be used to create **error correcting codes**.

Am I Ready for This Material

Essential to understanding the material of this section are a mastery of the following previously introduced concepts:

field

vector space over a field

linear combination of a sequence (or set) of vectors from a vector space

subspace of a vector space

span of a sequence of vectors from a vector space

spanning sequence of a vector space V

a sequence of vectors in a vector space is linearly dependent

a sequence of vectors in a vector space is linearly independent

basis of a subspace of a vector space

dimension of a vector space

finite field

Quiz

1. Solve the linear equation $2x + 1 = 2$ if the coefficients are in $\mathbb{F}_5$.

2. Solve the **linear system** with **augmented matrix** $\begin{pmatrix} 1 & 1 & 1 & 2 & 1 & | & 0 \\ 2 & 1 & 1 & 2 & 1 & | & 1 \\ 1 & 2 & 1 & 2 & 1 & | & 1 \end{pmatrix}$ if the entries are in $\mathbb{F}_3$.

3. Find a **basis** for the **null space** of the matrix $\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$ with entries in $\mathbb{F}_2$.

Quiz Solutions

New Concepts

We introduce a number of new concepts:

q-ary word of length n

weight of a q-ary word of length n

Hamming distance between two vectors

ball of radius r with center w

linear code of length n over $\mathbb{F}_q$

minimum weight of a linear code $\mathcal{C}$ of length n over $\mathbb{F}_q$,

generator matrix of a linear code of length n over $\mathbb{F}_q$

parity check matrix of a linear code of length n over $\mathbb{F}_q$

Theory (Why It Works)

Error correcting codes are used whenever a message is transmitted in a digital format over a "noisy" communication channel. This could be a phone call over a land line or wireless, email between two computers, a picture sent from outer space, an MP3 player interpreting digital music in a file, a computer memory system and many others. The "noise" could be human error, lightning, solar flares, imperfections in the equipment, deterioration in a computer memory device and so on, introducing errors by exchanging some of the digits of the message for other, incorrect digits.

The basic idea is to introduce redundancy into the message, enough, so that errors can be detected and corrected, but not so much that one has to send long messages relative to what we wish to transmit and consequently reducing the "information rate" and making transmission too costly.

A digital communication system begins with a message in the form of a string of symbols. This is input to an encoder which adds redundancy and creates a codeword. The codeword is sent over the communication channel where it may or may not be altered and out the other end comes a received string of symbols. This goes to a decoder which detects whether any errors have occurred. In a simple system, if an error has occurred then the sender is informed of this and asked to resend the message. In a more complicated scheme the decoder can correct errors as well as detect them and then sends the message on to the intended recipient. This is pictured schematically in **Figure** (5.8.1).

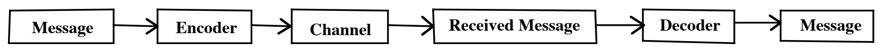


Figure 5.8.1: Sending a Message Over a Noisy Channel

Definition 5.29. By a ***message*** we will mean a string of symbols in some ***finite alphabet***. The message is ***binary*** if the alphabet has only two symbols. It is said to be ***q-ary***, with q some natural number, if the alphabet has q elements. Typically, the alphabet is taken to be a **finite field** $\mathbb{F}_q$ and consequently q is usually a power of a **prime number**.

In the typical situation the channel satisfies the following conditions:

1) The probability that a symbol α from the alphabet is transmitted and α is received is independent of α; and

2) The probability that a symbol α is sent and $\beta \neq \alpha$ is received is independent of α and β.

Suppose the alphabet has q symbols and the probability that α is sent and received is p. Then the probability that α is sent but α is not received is $1 - p$. Since there are $q-1$ possibilities for the received symbols and each is equally likely by assumption 2) it follows that the probability that α is sent and a fixed $\beta \neq \alpha$ is received is $\frac{1-p}{q-1}$.

It is also assumed that the channel, though noisy, is pretty good, meaning that p is close to one and, therefore, $1 - p$ is small.

Example 5.8.1. We want to send a message about how to move on a grid made up of horizontal and vertical lines. At any location one can go up, down, right or left.

In binary these can be encoded in the following way where we treat 0 and 1 as the elements of the **finite field** $\mathbb{F}_2$:

$$up = (0,0), down = (1,0), right = (0,1), left = (1,1) \tag{5.38}$$

These are the message digits that we wish to send but in the present form it is not useful since if an error occurs we cannot tell since it simply transforms one valid message into another valid message.

We can improve this by adding ***redundancy*** in the form of a ***check digit*** - adding a third digit to each message so that the number of one's is even, or the same thing, the sum of the digits is zero (remember our digits are elements of the field $\mathbb{F}_2$). With the introduction of this redundancy the expressions we use to communicate the directions become

$$up = (0,0,0), down = (1,0,1), right = (0,1,1), left = (1,1,0) \tag{5.39}$$

Now if a single error occurs it can be detected. This information could be communicated and a request made for resending the message, which is, of course, costly. So if we want to detect and correct errors, more redundancy is needed.

We can systematically add greater redundancy in the following way: If $\boldsymbol{w}$ is one of the four pairs of (5.38) follow $\boldsymbol{w}$ with a check digit then with $\boldsymbol{w}$ again. Thus,

$$(0,0) \to (0,0,0,0,0), (1,0) \to (1,0,1,1,0) \tag{5.40}$$

$$(0,1) \to (0,1,1,0,1), (1,1) \to (1,1,0,1,1) \tag{5.41}$$

Now if a single error occurs we can not only detect it but we can correct it by decoding the received vector as the one among the four vectors of (5.41) which is "closest" to it, in the sense that they differ in the minimum number of digits.

For example, if a received vector has a single one and four zeros then it differs from (0,0,0,0,0) in only one place but from all the others in two or more places. Therefore we would decode it as (0,0,0,0,0) = up.

We need to make the idea of **Example** (5.8.1) more precise and in order to do so we need to make some definitions.

Definition 5.30. Let $\mathbb{F}_q$ be a finite field. By a ***q-ary word of length n*** we will mean an element of the **vector space** $\mathbb{F}_q^n$ written as a row.

Definition 5.31. Let $\boldsymbol{x} = (a_1\ a_2 \dots\ a_n)$ be a **q-ary word of length n**. Then the ***weight*** of $\boldsymbol{x}$, $wt(\boldsymbol{x})$, is the number of i such that $a_i \neq 0$.

The **weight function** is used to introduce the notion of a distance between words, the ***Hamming distance***, introduced by the coding theory pioneer **Richard Hamming**.

Definition 5.32. Let $\boldsymbol{x} = (a_1\ a_2 \ldots\ a_n)$ and $\boldsymbol{y} = (b_1\ b_2 \ldots\ b_n)$ be two **q-ary words of length n**. Then the ***Hamming distance*** between $\boldsymbol{x}$ and $\boldsymbol{y}$, denoted by $d(\boldsymbol{x}, \boldsymbol{y})$, is the number of i such that $a_i \neq b_i$.

Notice that in the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ that $a_i \neq b_i$ if and only if the i^{th} component of $\boldsymbol{x} - \boldsymbol{y}$ is not equal to zero. This observation implies the following:

Theorem 5.8.1. *Let $\boldsymbol{x}, \boldsymbol{y}$ be words from $\mathbb{F}_q^n$. Then $d(\boldsymbol{x}, \boldsymbol{y}) = wt(\boldsymbol{x} - \boldsymbol{y})$. In particular, $d(\boldsymbol{x}, \boldsymbol{0}_n) = wt(\boldsymbol{x})$.*

This **distance function** satisfies the following:

Theorem 5.8.2. *1) For any vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_q^n, d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ with equality if and only if $\boldsymbol{x} = \boldsymbol{y}$.*

2) For vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{F}_q, d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$.

3) The "triangle inequalty holds": For vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{F}_q$ $d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$.

The first and second should be clear. The third is left as a **challenge exercise**.

An important concept, for both conceptual and theoretic purposes, is the notion of a ball of radius r about a vector $\boldsymbol{w}$.

Definition 5.33. Let $\boldsymbol{w}$ be a word in $\mathbb{F}_q^n$ and r a natural number. The ***ball of radius*** r ***with center*** $\boldsymbol{w}$, denoted by $B_r(\boldsymbol{w})$, consists of all the **q-ary words of length n** whose **Hamming distance** from $\boldsymbol{w}$ is less than or equal to r :

$$B_r(\boldsymbol{w}) = \{\boldsymbol{x} \in \mathbb{F}_q^n : d(\boldsymbol{w}, \boldsymbol{x}) \leq r\}.$$

Example 5.8.2. The **ball of radius one** with center at (0,0,0,0,0) consists of (0,0,0,0,0) and all the words of **weight** one. For $\boldsymbol{w} = (1, 0, 1, 1, 0), B_1(\boldsymbol{w}) =$

$$\{(1, 0, 1, 1, 0), (0, 0, 1, 1, 0), (1, 1, 1, 1, 0), (1, 0, 0, 1, 0), (1, 0, 1, 0, 0), (1, 0, 1, 1, 1)\}.$$

The **balls of radius one** centered at the four words shown in (5.40) and (5.41) do not intersect.

One can easily count the number of vectors in a **ball of radius r**. We state the result and leave it as a **challenge exercise**:

Theorem 5.8.3. *Let $\boldsymbol{w} \in \mathbb{F}_q^n$.*

1) Let t be a nonnegative integer. Then number of $\boldsymbol{x} \in \mathbb{F}_q^n$ such that $d(\boldsymbol{w}, \boldsymbol{x}) = t$ is $\binom{n}{t}(q-1)^t$.

2) Let r be a nonnegative integer. Then the number of vectors in $B_r(\boldsymbol{w})$ is

$$1 + n(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{r}(q-1)^r.$$

In **Theorem** (5.8.3) the expression $\binom{n}{t}$, which is read "n choose t" is the number of ways of picking t objects from a collection of n indistiguishable objects. It is equal to $\frac{n!}{t!(n-t)!}$.

We now give the most general definition of a code:

Definition 5.34. A ***code*** is a subset $\mathcal{C}$ of some finite **vector space** $\mathbb{F}_q^n$. The ***length*** of the code is n. If the number of elements in $\mathcal{C}$ is K then we say that $\mathcal{C}$ is an ***(n,K)-code*** over $\mathbb{F}_q$.

A code $\mathcal{C}$ of length n is said to be a ***linear code over*** $\mathbb{F}_q$ if $\mathcal{C}$ is a subspace of $\mathbb{F}_q^n$. If the dimension of $\mathcal{C}$ is k then we say that $\mathcal{C}$ is an ***(n,k) linear code*** over $\mathbb{F}_q$.

Example 5.8.3. The collection of four vectors in (5.40) and (5.41) is a (5,2) linear code over $\mathbb{F}_2$.

In **Example** (5.8.1) we added redundancy so that the **Hamming distance** between pairs of codes words were sufficiently great that we could detect and correct single errors. We now make this rigorous.

Definition 5.35. Let $\mathcal{C}$ be a code of length n over $\mathbb{F}_q$. The ***minimum distance*** of $\mathcal{C}$ is

$$d(\mathcal{C}) = min\{d(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{y}\}.$$

In other words it is the minimum distance obtained by any two distinct codewords from $\mathcal{C}$.

The importance of the **minimum distance** of a code is indicated by the following result:

Theorem 5.8.4. *Let $\mathcal{C}$ be an* **(n,K)- code** *over $\mathbb{F}_q$ and assume that $d(\mathcal{C}) = d$. Then the following hold:*

1) $\mathcal{C}$ can detect up to e errors as long as $d \geq e + 1$.

2) $\mathcal{C}$ can correct up to c errors as long as $d \geq 2c + 1$.

Conceptually, 1) holds because the **ball of radius d-1** centered at a codeword does not contain any other codewords. Also, 2) holds because the **balls of radius c** with $2c + 1 \leq d$ centered at the code words are disjoint.

Proof. 1) Suppose that a codeword $\boldsymbol{w}$ is transmitted and the word $\boldsymbol{x}$ is received and there are e errors with $e < d$. The number of errors is simply $d(\boldsymbol{w}, \boldsymbol{x})$. Since $d(\boldsymbol{w}, \boldsymbol{x}) = e < d$ it cannot be that $\boldsymbol{x}$ is another codeword and, consequently, we can detect that an error occured.

2) Suppose $\boldsymbol{w}$ is transmitted and $\boldsymbol{x}$ is received with c errors, where $2c + 1 \leq d$. We claim that for any codeword $\boldsymbol{w}' \neq \boldsymbol{w}$ that $d(\boldsymbol{x}, \boldsymbol{w}') > c$ and therefore amongst $\mathcal{C}, \boldsymbol{w}$ is the unique nearest neighbor to $\boldsymbol{x}$. To see this claim, assume to the contrary that $d(\boldsymbol{x}, \boldsymbol{w}') \leq c$ for some codeword $\boldsymbol{w}' \neq \boldsymbol{w}$.

$$d \leq d(\boldsymbol{w}, \boldsymbol{w}') \leq d(\boldsymbol{w}, \boldsymbol{x}) + d(\boldsymbol{x}, \boldsymbol{w}') \leq c + c = 2c < d \tag{5.42}$$

by the **triangle inequality** and the definition of d. We therefore have a contradiction. $\square$

There are several good reasons for working with **linear codes**. One is that they can be constructed using matrix multiplication. Another is that the computation of the **minimum distance** is simplified and does not require computing the **distances** between every pair of vectors in the code. Before showing this we require another definition.

Definition 5.36. Let $\mathcal{C}$ be an **(n,k)- linear code** over $\mathbb{F}_q$. The ***minimum weight*** of $\mathcal{C}$, denoted by $m(\mathcal{C})$ is

$$min\{wt(\boldsymbol{w}) : \boldsymbol{w} \in \mathcal{C}, \boldsymbol{w} \neq \mathbf{0}_n\}.$$

This next theorem indicates the relationship between $d(\mathcal{C})$ and $m(\mathcal{C})$ for a **linear code**.

Theorem 5.8.5. *Let $\mathcal{C}$ be an* **(n,k)- linear code** *over $\mathbb{F}_q$. Then $d(\mathcal{C}) = m(\mathcal{C})$.*

Proof. $d(\mathcal{C}) = min\{d(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{y}\}$.

However, $d(\boldsymbol{x}, \boldsymbol{y}) = wt(\boldsymbol{x} - \boldsymbol{y})$. Therefore $d(\mathcal{C}) = min\{wt(\boldsymbol{x} - \boldsymbol{y}) : \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{y}\}$.

Since $\mathcal{C}$ is a **linear code**, as $(\boldsymbol{x}, \boldsymbol{y})$ runs over all pairs from $\mathcal{C}$ with $\boldsymbol{x} \neq \boldsymbol{y}$, $\boldsymbol{x} - \boldsymbol{y}$ runs over all nonzero vectors in $\mathcal{C}$ and consequently, $min\{wt(\boldsymbol{x} - \boldsymbol{y}) : \boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{y}\} = m(\mathcal{C})$ as claimed. $\square$

As we shall see **linear codes** can be constructed with a designed **minimum weight** and in this way no computation is required to determine the **minimum distance** and therefore the error detecting and error correcting capacity of the code. In the next example we show how the code of (5.41) can be constructed from the original message by matrix multiplication.

Example 5.8.4. Let $G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$. Then

$$(0, 0)G = (0, 0, 0, 0, 0), (1, 0)G = (1, 0, 1, 1, 0)$$

$$(0, 1)G = (0, 1, 1, 0, 1), (1, 1)G = (1, 1, 0, 1, 1).$$

Notice that the sequence of rows of the matrix G of **Example** (5.8.4) is a **basis** for this **linear code**. This is an example of a generator matrix for a code.

Definition 5.37. Let $\mathcal{C}$ be an **(n,k)-linear code** over $\mathbb{F}_q$. Any matrix $k \times n$ matrix G whose rows consists of a **basis** for $\mathcal{C}$ is a ***generator matrix of*** $\mathcal{C}$. The matrix G is said to be ***systematic*** if G has the form $(I_k B)$ where B is a $k \times (n - k)$ matrix.

Note that since the rows of G are a **basis** that the **rank** of G is equal to k.

We can use a **generator matrix** to encode a message of length k by matrix multiplication: Given a message $\boldsymbol{m} = (a_1, a_1, \dots, a_k)$ encode this as $\boldsymbol{m}G$. If G is **systematic** then the first k digits of the codeword $\boldsymbol{m}G$ will be the message $\boldsymbol{m}$.

In addition to encoding messages on the transmission end we need a decoder on the receiving end to detect whether errors have occurred, correct them if possible and deliver the original message to the user. The parity check matrix will be essential to this purpose. However, before we introduce this we require some further definitions.

Definition 5.38. Let $\boldsymbol{x} = (a_1\ a_2 \dots\ a_n)$ and $\boldsymbol{y} = (b_1\ b_2 \dots\ b_n)$ be two vectors in $\mathbb{F}_q^n$. Then the ***dot product*** of $\boldsymbol{x}$ and $\boldsymbol{y}$, denoted by $\boldsymbol{x} \cdot \boldsymbol{y}$, is

$$a_1b_1 + a_2b_2 + \dots a_nb_n.$$

We will say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are ***orthogonal*** or ***perpendicular*** if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$.

The following summarizes many of the properties of the dot product:

Theorem 5.8.6. *Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be vectors in $\mathbb{F}_q^n$ and $c \in \mathbb{F}_q$. Then the following hold:*

1) $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$.

2) $\boldsymbol{x} \cdot [\boldsymbol{y} + \boldsymbol{z}] = \boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{x} \cdot \boldsymbol{z}$.

3) $c[\boldsymbol{x} \cdot \boldsymbol{y}] = (c\boldsymbol{x}) \cdot \boldsymbol{y} = \boldsymbol{x} \cdot (c\boldsymbol{y})$.

Proof. 1) This holds since the multiplication in $\mathbb{F}_q$ is commutative: If $\boldsymbol{x} = (x_1\ \dots\ x_n), \boldsymbol{y} = (y_1\ \dots\ y_n)$ then for each $i, x_iy_i = y_ix_i$.

2) This holds since the distributive property holds in $\mathbb{F}_q$: If also $\boldsymbol{z} = (z_1\ \dots\ z_n)$ then for each i we have

$$x_i(y_i + z_i) = x_iy_i + x_iz_i.$$

3) This holds because the multiplication in $\mathbb{F}_q$ is associative and commutative: For each i

$$c(x_iy_i) = (cx_i)y_i = (x_ic)y_i = x_i(cy_i).$$

$\square$

So, in some ways the dot product in $\mathbb{F}_q^n$ is similar to the **dot product** in $\mathbb{R}^n$; however, it is extremely important to know that the dot product on the vector space $\mathbb{F}_q^n$ differs in a very significant way: In $\mathbb{R}^n$ the only vector $\boldsymbol{x}$ for which $\boldsymbol{x} \cdot \boldsymbol{x} = 0$ is the zero vector. This may not be the case over finite fields. For example, if (1,2) is in $\mathbb{F}_5^2$ then

$$(1, 2) \cdot (1, 2) = 1^2 + 2^2 = 5 = 0$$

in $\mathbb{F}_5$.

We will also need the notion of the orthogonal complement to a vector or a **subspace** of $\mathbb{F}_q^n$ and we now introduce this concept.

Definition 5.39. Let $\mathcal{C}$ be a **subspace** of $\mathbb{F}_q^n$. The ***orthogonal complement*** to $\mathcal{C}$, denoted by $\mathcal{C}^\perp$, consists of all those vectors which are **orthogonal** to every vector in $\mathcal{C}$:

$$\mathcal{C}^\perp = \{\boldsymbol{x} \in \mathbb{F}_q^n : \boldsymbol{x} \cdot \boldsymbol{w} = 0 \text{ for all } \boldsymbol{w} \in \mathcal{C}\}.$$

Assume that $\mathcal{C}$ is a **subspace** of $\mathbb{F}_q^n$ of **dimension** k. When we are considering $\mathcal{C}$ to be a **linear code** over $\mathbb{F}_q$ then we often refer to $\mathcal{C}^\perp$ as the ***dual code of*** $\mathcal{C}$.

Theorem 5.8.7. *Assume $\mathcal{C}$ is an* **(n,k) linear code** *over $\mathbb{F}_q$. Then the* **dual code** *$\mathcal{C}^\perp$ is an (n,n-k) linear code.*

Proof, We need to know that $\mathcal{C}^\perp$ is a **subspace** of $\mathbb{F}_q^n$ and has **dimension** $n-k$. Since $\mathcal{C}$ has **dimension** k there exists a **basis** $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ of $\mathcal{C}$. We claim that $\boldsymbol{y} \in \mathcal{C}^\perp$ if and only if $\boldsymbol{x}_i \cdot \boldsymbol{y} = 0$ for all $i, 1 \le i \le k$. If $\boldsymbol{y} \in \mathcal{C}^\perp$ then $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ for every $\boldsymbol{x} \in \mathcal{C}$ so, in particular, $\boldsymbol{x}_i \cdot \boldsymbol{y} = 0$. Therefore to prove our claim we need to prove if $\boldsymbol{x}_i \cdot \boldsymbol{y} = 0$ for every i then $\boldsymbol{y} \in \mathcal{C}^\perp$, that is, we need to show for an arbitrary vector $\boldsymbol{x} \in \mathcal{C}$ that $\boldsymbol{x} \cdot \boldsymbol{y} = 0$.

Since $\boldsymbol{x} \in \mathcal{C}$ and $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ is a **basis** of $\mathcal{C}$ it follows that there are scalars $c_1, \dots, c_k \in \mathbb{F}_q$ such that $\boldsymbol{x} = c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k$. Then by **Theorem** (5.8.6) it follows that

$$\boldsymbol{x} \cdot \boldsymbol{y} = [c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k] \cdot \boldsymbol{y} =$$

$$c_1(\boldsymbol{x}_i \cdot \boldsymbol{y}) + \cdots + c_k(\boldsymbol{x}_k \cdot \boldsymbol{y}) = 0.$$

Now let A be the matrix whose i^{th} row is $\boldsymbol{x}_i$. It then follows that $\boldsymbol{y} \in \mathcal{C}^\perp$ if and only if $\boldsymbol{y}^{Tr}$ is in the **null space** of A. Since the latter is a **subspace** of the space $\mathbb{F}_q^n$ of column vectors and since the **transpose** satisfies $(\boldsymbol{y}+\boldsymbol{z})^{Tr} = \boldsymbol{y}^{Tr} + \boldsymbol{z}^{Tr}$ and $(c\boldsymbol{y})^{Tr} = c\boldsymbol{y}^{Tr}$ it follows that $\mathcal{C}^{Tr}$ is a **subspace** of the row space $\mathbb{F}_q^n$. Moreover, by **Theorem** (5.5.4) it also follows that $dim(\mathcal{C}) + dim(\mathcal{C}^\perp) = n$ so that $dim(\mathcal{C}^\perp) = n - k$. $\square$

Example 5.8.5. For the code $\mathcal{C} = Span\,((1,0,1,1,0),(0,1,1,0,1))$ the **dual code** is $Span\,((1,0,0,1,0),(0,1,0,0,1),(1,1,1,0,0))$ which consists of the eight vectors

$$(0,0,0,0,0),(1,0,0,1,0),(0,1,0,0,1),(1,1,1,0,0)$$

$$(1,1,0,1,1),(0,1,1,1,0),(1,0,1,0,1),(0,0,1,1,1)$$

The **dual code** is what we need to define a parity check matrix for an **(n,k) linear code** $\mathcal{C}$ over $\mathbb{F}_q$.

Definition 5.40. Let $\mathcal{C}$ be an **(n,k) linear code** over $\mathbb{F}_q$. Any **generator matrix** H for the dual code $\mathcal{C}^{\perp}$ of $\mathcal{C}$ is a ***parity check matrix for*** $\mathcal{C}$.

Example 5.8.6. From **Example** (5.8.5) the matrix

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

is a **parity check matrix** for the binary code

$$\mathcal{C} = \{(0,0,0,0,0),(1,0,1,1,0),(0,1,1,0,1),(1,1,0,1,1)\}.$$

In part, the importance of the **parity check matrix** is indicated by the following:

Theorem 5.8.8. *Let $\mathcal{C}$ be an* **(n,k) linear code** *over $\mathbb{F}_q$ and H a* **parity check matrix**. *Then $\boldsymbol{w} \in \mathbb{F}_q^n$ is a codeword if and only if $H\boldsymbol{w}^{Tr} = \mathbf{0}_{n-k}$.*

Proof. Suppose $\boldsymbol{w} \in \mathcal{C}$. Then $\boldsymbol{w}$ is perpendicular to every row of H by the definition of H. In particular, the product of $\boldsymbol{w}^{Tr}$ with each row of H^{Tr} is zero and therefore $H\boldsymbol{w}^{Tr} = \mathbf{0}_{n-k}$.

Conversely, the **rank** of H is $n-k$ since its rows are **linearly independent**. Therefore the **null space** of H has dimension $n-(n-k) = k$. However, $null(H)$ contains $\{\boldsymbol{w}^{Tr} : \boldsymbol{w} \in \mathcal{C}\}$ which has **dimension** k and therefore this is all of $null(H)$. $\square$

Given a **systematic generator matrix** $G = (I_k B)$ for a **linear code** $\mathcal{C}$ it is easy to obtain a **parity check matrix** and this is the subject of the next theorem.

Theorem 5.8.9. *Assume that $G = (I_k B)$ is a* **systematic generator matrix** *for an* **(n,k) linear code** *$\mathcal{C}$ over $\mathbb{F}_q$. Then $H = (-B^{Tr} I_{n-k})$ is a* **parity check matrix** *for $\mathcal{C}$.*

Proof. H is an $(n-k) \times n$ matrix. Since the last $n-k$ columns are a **basis** for $\mathbb{F}_q^{n-k}$, H has **rank** $n-k$. By **Theorem** (5.8.8) we will be done if we can show that $G^{Tr}H = \mathbf{0}_{(n-k)\times k}$.

$$HG^{Tr} = (-B^{Tr} I_{n-k}) \begin{pmatrix} I_k \\ B^{Tr} \end{pmatrix} = -B^{Tr} I_k + I_{n-k} B^{Tr} = -B^{Tr} + B^{Tr} = \mathbf{0}_{(n-k)\times k}.$$

$\square$

Example 5.8.7. The matrix $G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}$ is a **systematic generator matrix** for the code $\mathcal{C} = Span\,((1,0,1,1,0),(0,1,1,0,1))$. The parity check matrix we obtain from this is $H' = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$

For a **linear code** $\mathcal{C}$, a parity check matrix H can be used to determine the **minimum weight** of $\mathcal{C}$ as we prove in the next theorem.

Theorem 5.8.10. *Let H be a* **parity check matrix** *for a* **linear (n,k) code** *$\mathcal{C}$ over $\mathbb{F}_q$. Assume that every sequence of $d-1$ columns of H is* **linearly independent** *but some sequence of d columns is* **linearly dependent**. *Then $m(\mathcal{C}) = d$.*

Proof. Suppose for the sake of the proof that the sequence of the first d columns, $S = (\boldsymbol{c}_1, \dots, \boldsymbol{c}_d)$ of H are **linearly dependent**. Let

$$a_1\boldsymbol{c}_1 + a_2\boldsymbol{c}_2 + \cdots + a_d\boldsymbol{c}_d = \mathbf{0}_n$$

be a **non-trivial dependence relation** of S. Then the vector $\boldsymbol{x} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ satisfies $H\boldsymbol{x} = \mathbf{0}_{n-k}$ and therefore $\boldsymbol{w} = \boldsymbol{x}^{Tr} \in \mathcal{C}$ by Theorem (5.8.8). Since $wt(\boldsymbol{w}) = d$ we conclude that $m(\mathcal{C}) \leq d$.

On the other hand suppose $\boldsymbol{y}$ is a vector with **weight** less than d and $H\boldsymbol{y}^{Tr} = \mathbf{0}_{n-k}$. Suppose $\boldsymbol{y}$ is nonzero and let the nonzero entries in $\boldsymbol{y}$ be $b_{i_1}, b_{i_2}, \dots, b_{i_t}$ where $t < d$. Since $H\boldsymbol{y}^{Tr} = \mathbf{0}_{n-k}$ we conclude that

$$b_{i_1}\boldsymbol{c}_{i_1} + b_{i_2}\boldsymbol{c}_{i_2} + \cdots + b_{i_t}\boldsymbol{c}_{i_t} = \mathbf{0}_{n-k}$$

However, this implies that the sequence of columns $(\boldsymbol{c}_{i_1}, \boldsymbol{c}_{i_2}, \ldots, \boldsymbol{c}_{i_t})$ is **linearly dependent**. Since $t < d$ this contradicts our hypothesis. Thus, no nonzero vector in $\mathcal{C}$ has **weight** less than d and the **minimum weight** of $\mathcal{C}$ is exactly d. □

It follows from **Theorem** (5.8.10) if the columns of a **parity check matrix** H of a **linear code** $\mathcal{C}$ are all distinct then the **minimum weight** is at least two and we can always detect a single error. If no two columns of H are multiples, equivalently, pairs of columns of H are **linearly independent** then the **minimum weight** of the code is at least three and we can correct single errors. Note for binary codes a pair of nonzero vectors is **linearly independent** if and only if they are distinct.

Example 5.8.8. Let H be the matrix whose columns are all the nonzero vectors in $\mathbb{F}_2^3$. We use H as the parity check matrix of a code. We will treat the vectors in $\mathbb{F}_2^3$ as a binary expression for a natural number where $\begin{pmatrix}1\\0\\0\end{pmatrix} = 1, \begin{pmatrix}0\\1\\0\end{pmatrix} = 2, \begin{pmatrix}0\\0\\1\end{pmatrix} = 4.$ This will be of use in our decoding scheme. We order the columns from 1 to 7. Thus,

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

Since the sequence of standard basis vectors of $\mathbb{F}_2^3$ is a subsequence of the columns of H it follows that H has **rank** three. Let $\mathcal{H}(3,2)$ denote the code that is dual to the row space of H. In this notation the (3,2) indicates that the columns of the parity check matrix H are the 3-vectors over $\mathbb{F}_2$. This resulting code is referred to as a **binary Hamming code**. This is a **linear (7,4) code** over $\mathbb{F}_2$.

Since the columns of H are all distinct and this is a binary matrix, the **minimum weight** of $\mathcal{H}(3,2)$ is at least 3. On the other hand the sum of the first three columns of H is the zero vector and therefore the **minimum weight** is exactly 3. Thus, $\mathcal{H}(3,2)$ is a 1-error correcting code.

Notice that a **ball of radius one** centered at a word contains 1 + 7 = 8 words. If we consider the balls of radius one around the 16 code words then these are disjoint and so the number of words they jointly cover is $16 \times 8 = 128 = 2^7$. That, is, each word is contained in exactly one of these balls.

Let $\boldsymbol{e}_i, i = 1, 2, \ldots, 7$ denote the standard basis of $\mathbb{F}_2^7$. Now suppose some codeword $\boldsymbol{w}$ is sent and $\boldsymbol{x}$ is received and one error occured, say in the i^{th} position. Then by the definition of $\boldsymbol{e}_i, \boldsymbol{x} = \boldsymbol{w} + \boldsymbol{e}_i$. We can deduce that an error has occurred since $H\boldsymbol{x}^{Tr} \neq \mathbf{0}_3$. But we get more information. The nonzero vector $H\boldsymbol{x}^{Tr}$ is called the ***syndrome*** of $\boldsymbol{x}$, and is denoted by $S(\boldsymbol{x})$. In this example, it will tell us precisely where the error occurred.

Since $\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{e}_i, S(\boldsymbol{x}) = H\boldsymbol{x}^{Tr} = H(\boldsymbol{w}^{Tr} + \boldsymbol{e}_i^{Tr}) = H\boldsymbol{w}^{Tr} + H\boldsymbol{e}_i^{Tr} = H\boldsymbol{e}_i^{Tr}$ because $\boldsymbol{w}$ is in the code and therefore $H\boldsymbol{w}^{Tr} = \mathbf{0}_3$.

Since $\boldsymbol{e}_i$ is the i^{th} standard basis vector of $\mathbb{F}_2^7$, $H\boldsymbol{e}_i^{Tr}$ is the i^{th} column of H. This gives us a decoding scheme:

Take the received word $\boldsymbol{x}$ and compute its syndrome $S(\boldsymbol{x}) = H\boldsymbol{x}^{Tr}$. If $S(\boldsymbol{x}) = \mathbf{0}_3$ then $\boldsymbol{x}$ is codeword and the intended message can be obtained from the received word $\boldsymbol{x}$ (though how depends on the encoder used). If $S(\boldsymbol{x}) = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \neq \mathbf{0}_3$ then let i be the natural number with binary expansion $a_3a_2a_1$. Set $\boldsymbol{w} = \boldsymbol{x} + \boldsymbol{e}_i$. This will be a codeword (the unique one at distance one from $\boldsymbol{x}$) and decode as $\boldsymbol{w}$.

As a concrete example, suppose the word $\boldsymbol{x} = (1, 1, 0, 0, 0, 0, 1)$ is received. Then the syndrome of this vector is

$$S(\boldsymbol{x}) = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

The vector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is binary for 4. Thus, if one error occurred it was in the fourth position and the codeword sent was (1,1,0,1,0,0,1).

The code of **Example** (5.8.8) is one in a family of 1-error correcting codes where the **balls of radius one** centered at the codewords cover all the words. Such codes are said to be ***perfect 1-error correcting codes***

Clearly, one needs to do better than be able to correct one error and it is not difficult to define such codes using **Theorem** (5.8.10). We show how to construct **linear codes** with a designed **minimum weight**.

BCH-codes

Let $\alpha_1, \alpha_2, \dots, \alpha_{q-1}$ be the nonzero elements of the finite field $\mathbb{F}_q$ and let t be a natural number, $t \leq q-1$.

Let H be the following matrix

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \alpha_1 & \alpha_2 & \dots & \alpha_{q-1} \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_{q-1}^2 \\ \vdots & \vdots & \dots & \vdots \\ \alpha_1^{t-1} & \alpha_2^{t-1} & \dots & \alpha_{q-1}^{t-1} \end{pmatrix}$$

We will show that any t columns from H are linearly independent. Suppose $\beta_1, \beta_2, \dots, \beta_t$ is a subset of $\{\alpha_1, \alpha_2, \dots, \alpha_{q-1}\}$. Consider the square matrix made from the column $\begin{pmatrix} 1 \\ \beta_i \\ \beta_i^2 \\ \vdots \\ \beta_i^{t-1} \end{pmatrix}$. This matrix is $\begin{pmatrix} 1 & 1 & \dots & 1 \\ \beta_1 & \beta_2 & \dots & \beta_t \\ \beta_1^2 & \beta_2^2 & \dots & \beta_t^2 \\ \vdots & \vdots & \dots & \vdots \\ \beta_1^{t-1} & \beta_2^{t-1} & \dots & \beta_t^{t-1} \end{pmatrix}$.

The **determinant** of this matrix is $\Pi_{1 \leq i < j \leq t}(\beta_j - \beta_i)$. Since for $i \neq j, \beta_i \neq \beta_j$, we conclude that this matrix has nonzero determinant and is **invertible**. Therefore any sequence of t columns is **linearly independent**. Since there are only t rows the **rank** of H is exactly t and a sequence of any $t+1$ columns is **linearly dependent**.

From what we have shown, if $\mathcal{C}$ is the code consisting of all words $\boldsymbol{w}$ in $\mathbb{F}_q^{q-1}$ which satisfy $H\boldsymbol{w}^{Tr} = \mathbf{0}_t$ then $\mathcal{C}$ is a **(q-1, q-t-1) linear code** with **minimum weight** $t+1$. Therefore, if $2e+1 \leq t$ this code can be used to correct e errors.

To be practical there needs to be some algorithm to do the encoding and decoding. Any **generator matrix** can be used for the encoding. An algorithm does exist for these codes, based on ideas from number theory developed by the Indian mathematician **Ramanujan**. The codes are known as **BCH codes**. They were invented in 1959 by Hocquenghem and independently by Bose and Ray-Chaudhuri and they have been used fairly extensively.

Example 5.8.9. The matrix $H = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 2 & 2 & 4 & 1 \\ 1 & 1 & 6 & 1 & 6 & 6 \end{pmatrix}$ is the **parity check matrix** for the **BCH code of length 6**, with designed **minimum weight** 5 over the field $\mathbb{F}_7$

which we denote by $\mathcal{BCH}(7,4)$. The seven indicates that we are using the field of 7 elements and that we are using just enough powers of the elements of the field to make 4 rows (and hence a matrix of **rank** 4). Since there are 6 columns and the **rank** is four, the **nullity** is two and therefore the code $\mathcal{BCH}(7,4)$ has **dimension** 2 and is therefore a **(6,2) code** over $\mathbb{F}_7$ with **minimum weight** 5. It is a double error correcting code.

Exercises

In exercises 1 - 4 find the **Hamming distance** between the two given binary words:

1. (000111), (110100)

2. (101010), (010101)

3. (111100), (101001)

4. (000001), (011010)

In exercises 5 and 6 find the **Hamming distance** between the two given words from $\mathbb{F}_5^6$.

5. (121314), (020124)

6. (001234), (321014)

7. Let $\mathcal{C}$ consist of the following five binary vectors:

$$(00000), (01111), (11100), (10011), (11010)$$

Determine the **minimum distance** $d(\mathcal{C})$ of this code.

8. In $\mathbb{F}_3^3$ find all vectors $\boldsymbol{x}$ such that $\boldsymbol{x} \cdot \boldsymbol{x} = 0$.

9. In $\mathbb{F}_5^5$ find all vectors $\boldsymbol{x}$ such that $\boldsymbol{x} \cdot \boldsymbol{x} = 0$.

10. Define a binary code as follows: given a message abc add three check digits xyz so that the code word becomes $abcxyz$ where $x = a+b, y = a+c, z = b+c$ (this is a linear code).

a) Write down a **systematic generator matrix** for this code.

b) Write down a **parity check matrix** for this code.

c) Determine the **minimum distance** of this code which is equal to the **minimum weight** of the code.

11. Extend the code of exercise 9 by adding another parity check, w where $w = a+b+c$. Determine the parameters of this code (length, dimension, minimum distance).

In exercises 12 and 13 the given matrix is the **generator matrix** for a code $\mathcal{C}$ over the given field. Determine if this code has a **systematic generator matrix** and, if so, find such a matrix.

12. $G = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$ over $\mathbb{F}_2$.

13. $G = \begin{pmatrix} 1 & 1 & 1 & 2 & 1 & 2 & 1 \\ 1 & 2 & 2 & 0 & 2 & 0 & 1 \\ 1 & 1 & 2 & 0 & 0 & 1 & 0 \end{pmatrix}$ over $\mathbb{F}_3$.

14. Find a **parity check matrix** for the code $\mathcal{C}$ of exercise 11.

15. Find a **parity check matrix** for the code $\mathcal{C}$ of exercise 12.

16. Find the parameters of the code $\mathcal{C}$ of exercise 11.

17. Find the parameters of the code $\mathcal{C}$ of exercise 12.

Let $H = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 & 1 & 1 & 2 \end{pmatrix}$ be the **parity check matrix** of a code $\mathcal{C}$ over $\mathbb{F}_3$. In exercises 16 - 18 determine if the given vector is in $\mathcal{C}$.

18. (1202200)

19. (2010012)

20. (1121121)

Let $H = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$ be the **parity check matrix** of the Hamming code $\mathcal{H}(3, 2)$. In exercises 19 - 22 compute the **syndrome** of the given vector and determine if it is a code word. If it is not a code work, assuming one error occurred, correct this error.

21. (0110100)

22. (1110100)

23. (1111111)

24. (1101100)

Let H be the **parity check matrix** of the code $\mathcal{BCH}(7, 4)$ of **Example** (5.8.9). In exercises 25 - 28 determine if the given word is in code.

25. (136310)

26. (365621)

27. (155102)

28. (211406)

Challenge Exercises (Problems)

A **linear code** $\mathcal{C}$ is called **self-dual** if it is contained in the **dual code** $\mathcal{C}^{\perp}$, equivalently, if every pair of code words in $\mathcal{C}$ are orthogonal.

1. Assume that $\mathcal{C}$ is a self-dual **(2n,k) code** over the field $\mathbb{F}_q$. Prove that $k \leq n$.

2. a) Show that in the **binary Hamming code** $\mathcal{H}(3,2)$ there are equally many code-words of **weight** w and $7 - w$.

b) Without writing out all the code words, prove that there are 7 code words of weight 3 and 7 of weight 4 in the **binary Hamming code** $\mathcal{H}(3,2)$. (Hint: Make use of a) and the fact that the **minimum weight** of $\mathcal{H}(3,2)$ is 3.)

3. Let $\overline{\mathcal{H}}(3,2)$ be the ***extended binary Hamming code***, that is, the code obtained from $\mathcal{H}(3,2)$ by adding an overall parity check. Prove that this code contains the zero vector, the all one vector and 14 vectors of weight 4. (Hint: Make use of 2b)).

4. Let $\boldsymbol{x}$ be a word in $\mathbb{F}_2^n$. Prove that $\boldsymbol{x} \cdot \boldsymbol{x} = 0$ if and only if the **weight** of $\boldsymbol{x}$ is even.

For a vector $\boldsymbol{x} = (x_1\ x_2\ \ldots\ x_n)$ in $\mathbb{F}_2^n$ let the ***support*** of $\boldsymbol{x}$, $spt(\boldsymbol{x})$, be the subset of $\{1, 2, \ldots, n\}$ such that $x_i \neq 0$. For example, the support of (1001011) is $\{1, 4, 6, 7\}$.

5. Let $\boldsymbol{x}, \boldsymbol{y}$ be words in $\mathbb{F}_2^n$. Prove that $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ if and only if there are an even number of elements in the intersection of $spt(\boldsymbol{x}) \cap spt(\boldsymbol{y})$.

6. Prove that the extended binary Hamming code $\overline{\mathcal{H}}(3,2)$ is a self-dual code. (Hint: Use 4 and 5).

7. Suppose $\mathcal{C}$ is a **(23, 12) binary linear code** and the **minimum distance** is 7. Prove that the **balls of radius 5** centered at the code words are disjoint and cover all the vectors in $\mathbb{F}_2^{23}$. (This means that this is a **perfect** 3-error correcting code. Such a code exists and is unique. It is known as the ***binary Golay code***).

8. Let $H = (\boldsymbol{c}_1\ \boldsymbol{c}_2\ \ldots\ \boldsymbol{c}_n)$ be a **parity check matrix** of a code $\mathcal{C}$ over the field $\mathbb{F}_q$. Let $\alpha_i \in \mathbb{F}_q, \alpha \neq 0, i = 1, 2, \ldots, n$ and set $H' = (\alpha_1\boldsymbol{c}_1\ \alpha_2\boldsymbol{c}_2\ \ldots\ \alpha_n\boldsymbol{c}_n)$ be the matrix obtained from H by multiplying the column $\boldsymbol{c}_i$ by α_i. Let $\mathcal{C}'$ be the code with **parity check matrix** H'. Prove that the parameters (length, dimension, minimum distance, number of words of a given weight) of these two codes are identical.

9. Let $\pi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ be a permutation, that is, a one-to-one (and onto) function. Let $H = (\boldsymbol{c}_1\ \boldsymbol{c}_2\ \ldots\ \boldsymbol{c}_n)$ be a **parity check matrix** of a code $\mathcal{C}$ over the field $\mathbb{F}_q$ and set $H' = (\boldsymbol{c}_{\pi(1)}\ \boldsymbol{c}_{\pi(2)}\ \ldots\ \boldsymbol{c}_{\pi(n)})$ be the matrix obtained from H by

permuting the columns and $\mathcal{C}'$ the code with **parity check matrix** H'. Prove that the codes $\mathcal{C}$ and $\mathcal{C}'$ have identical parameters.

Codes related as in challenge exercises 8 and 9 are said to be **equivalent**.

10. Prove part 3) of **Theorem** (5.8.2).

11. Prove **Theorem** (5.8.3).

Quiz Solutions

1. $x = 3$.

2. $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}$.

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

Chapter 6

Linear Transformations

6.1. Introduction to Linear Transformations on Abstract Vector Spaces

The notion of a linear transformation on abstract vector spaces is introduced and we establish a number of important properties.

Am I Ready for This Material

Readiness Quiz

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

The following are those concepts that will be referred to in this module and which you should have mastered in order to achieve full understanding.

matrix

null space of a matrix

product of an $m \times n$ matrix A and an $n-$vector x

span of a sequence of vectors

function, transformation, map

domain of a transformation

codomain of a transformation

image of an element under a transformation

range of a transformation

linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$

standard matrix of a linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$

basis of a vector space

coordinate vector of v with respect to a basis

You will also need to know several of the methods and procedures that we have developed, especially **Gaussian elimination**.

Quiz

In 1 and 2 find the **domain and codomain** of the given **tranformation**.

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

2. $T(a + bx + cx^2) = ax + \frac{b}{2}x^2 + \frac{c}{3}x^3$.

In 3 and 4 show that the given **transformation** from $\mathbb{R}^2$ to $\mathbb{R}^3$ is not a **matrix transformation**.

3. $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x+3y \\ 3x+4y-1 \\ x+y \end{pmatrix}$

4. $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ xy \end{pmatrix}$

5. Find the **standard matrix** of the following transformation:

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+2y+3z \\ 2x+y-4z \\ 3x+7y-z \end{pmatrix}$$

6. Let $\mathcal{B} =$

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

a) Verify that $\mathcal{B}$ is a **basis** for $M_{2\times 2}(\mathbb{R})$; and

b) Determine the **coordinate vector** for each **standard basis vector** of $M_{2\times 2}(\mathbb{R})$ with respect to $\mathcal{B}$.

Quiz Solutions

New Concepts

In this section we introduce three new concepts:

The function $T : V \to W$ is a **linear transformation**
The **vector space** V is the **direct sum of subspaces X and Y**

The **projection map** $Proj_{(X,Y)}$

Theory (Why It Works)

In section (3.1), in connection with defining the **product of an $m \times n$ matrix and an n-vector**, we introduced the concept of a **linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$**. Recall, this means that T is a **function** from $\mathbb{R}^n$ to $\mathbb{R}^m$ which satisfies the following two conditions:

1) For all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n, T(\boldsymbol{u}+\boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$; and
2) For any $\boldsymbol{u} \in \mathbb{R}^n$ and scalar $c, T(c\boldsymbol{u}) = cT(\boldsymbol{u})$.

Further recall, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** and $\boldsymbol{a}_j$ is the **image** of $\boldsymbol{e}_j^n$, the j^{th} **standard basis vector** of $\mathbb{R}^n$, then the matrix A whose columns are $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ is the **standard matrix** of T. We went on to prove if $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is an arbitrary **n-vector** then $T(\boldsymbol{x}) = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$ and we defined the **product** $A\boldsymbol{x}$ of A and $\boldsymbol{x}$ to be equal to $x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n = T(\boldsymbol{x})$.

There is nothing particularly special about the vector spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, in fact, linear transformations can be defined between any two **vector spaces** over the same **field** $\mathbb{F}$. This is the content of our next definition.

Definition 6.1. Assume V and W are **vector spaces** over the **field** $\mathbb{F}$. A **function** $T : V \to W$ is a ***linear transformation*** if the following hold:

1) For all $\boldsymbol{u}, \boldsymbol{v} \in V, T(\boldsymbol{u}+\boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$; and
2) For any $\boldsymbol{u} \in V$ and scalar $c, T(c\boldsymbol{u}) = cT(\boldsymbol{u})$.

Example 6.1.1. 1. Let $V = M_{2\times 2}(\mathbb{R}), W = \mathbb{R}$ and define T by $T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a + c$. This is called the ***trace map***.

2. Define $D : \mathbb{R}[x] \to \mathbb{R}[x]$ by $D(a_0 + a_1x + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1}$.

3. Let $A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 0 & -1 \end{pmatrix}$. Define $T : M_{34} \to M_{24}$ by $T(\boldsymbol{m}) = A\boldsymbol{m}$.

4. Let F be the collection of functions from $\mathbb{R}$ to $\mathbb{R}$ and $a \in \mathbb{R}$. Define $E_a : F \to \mathbb{R}$ by $E_a(f) = f(a)$. This is called ***evaluation at*** a.

For f and g in F we have $E_a(f+g) = (f+g)(a) = f(a) + g(a) = E_a(f) + E_a(g)$.

Also, for $f \in F$ and c a scalar, $E_a(cf) = (cf)(a) = cf(a) = cE_a(f)$.

Thus, $E_a : F \to \mathbb{R}$ is a **linear transformation**.

The following theorem reduces the criteria for a function T between two vectors spaces to a single condition

Theorem 6.1.1. *Let $T : V \to W$ be a **function**. Then T is a linear transformation if and only if for every pair of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ and scalars $c_1, c_2 \in \mathbb{R}$, $T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2)$.*

Proof. Suppose T is a linear transformation, $\boldsymbol{v}_1, \boldsymbol{v}_2$ are in V and $c_1, c_2 \in \mathbb{R}$. Then $T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = T(c_1\boldsymbol{v}_1) + T(c_2\boldsymbol{v}_2)$ by the additive condition for a linear transformation. However, $T(c_1\boldsymbol{v}_1) = c_1T(\boldsymbol{v}_1), T(c_2\boldsymbol{v}_2) = c_2T(\boldsymbol{v}_2)$ by the scalar condition for a linear transformation, from which it follows that

$$T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2).$$

On the other hand, suppose T satisfies the given property. Then when we take $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V, c_1 = c_2 = 1$ we get $T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = T(\boldsymbol{v}_1) + T(\boldsymbol{v}_2)$ which is the first condition required of a **linear transformation**.

Taking $\boldsymbol{v}_1 = \boldsymbol{v}, \boldsymbol{v}_2 = \boldsymbol{0}, c_1 = c, c_2 = 0$ we get $T(c\boldsymbol{v}) = cT(\boldsymbol{v})$, which is the second condition. □

Theorem 6.1.2. *Let V be a* **vector space** *and assume that X and Y are* **subspaces** *of V which satisfy:*

$X \cap Y = \{\boldsymbol{0}_V\}$ and $X + Y = V$. Then the following hold:

1) For every vector $\boldsymbol{v} \in V$ there are unique vectors $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$ such that $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y}$.

2) If $\boldsymbol{x} \in X, \boldsymbol{y} \in Y$ such that $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y}$ then set $Proj_{(X,Y)}(\boldsymbol{v}) = \boldsymbol{x}$. Then $Proj_{(X,Y)}$ is a linear transformation from V to V.

Proof. 1) Recall that $X + Y$ consists of all vectors which can be written in the form $\boldsymbol{x} + \boldsymbol{y}$ where $\boldsymbol{x}$ is from X and $\boldsymbol{y}$ is from Y. Since we are assuming that $X + Y = V$, we are assuming that every vector can be written this way and so we only need to show that any such expression is unique.

Assume that

$$\boldsymbol{v} = \boldsymbol{x}_1 + \boldsymbol{y}_1 = \boldsymbol{x}_2 + \boldsymbol{y}_2. \tag{6.1}$$

We need to show that $\boldsymbol{x}_1 = \boldsymbol{x}_2, \boldsymbol{y}_1 = \boldsymbol{y}_2$.

Subtract $\boldsymbol{x}_2 + \boldsymbol{y}_1$ from both sides of (6.1). We obtain

$$\boldsymbol{x}_1 - \boldsymbol{x}_2 = \boldsymbol{y}_2 - \boldsymbol{y}_1 \tag{6.2}$$

The left hand side of (6.2) is in X and the right hand side is in Y. Since they are equal they belong to $X \cap Y = \{\boldsymbol{0}_V\}$. Therefore $\boldsymbol{x}_1 - \boldsymbol{x}_2 = \boldsymbol{y}_2 - \boldsymbol{y}_1 = \boldsymbol{0}_V$ from which we conclude that $\boldsymbol{x}_1 = \boldsymbol{x}_2, \boldsymbol{y}_1 = \boldsymbol{y}_2$ are required.

2) Suppose $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ and c_1, c_2 are scalars. We need to show that $Proj_{(X,Y)}(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1Proj_{(X,Y)}(\boldsymbol{v}_1) + c_2Proj_{(X,Y)}(\boldsymbol{v}_2)$.

Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in Y$ such that

$$\boldsymbol{v}_1 = \boldsymbol{x}_1 + \boldsymbol{y}_1, \boldsymbol{v}_2 = \boldsymbol{x}_2 + \boldsymbol{y}_2 \tag{6.3}$$

Then, by the definition of $Proj_{(X,Y)}$ we have

$$Proj_{(X,Y)}(\boldsymbol{v}_1) = \boldsymbol{x}_1, Proj_{(X,Y)}(\boldsymbol{v}_2) = \boldsymbol{x}_2 \tag{6.4}$$

By (6.3) we have

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 = c_1(\boldsymbol{x}_1 + \boldsymbol{y}_1) + c_2(\boldsymbol{x}_2 + \boldsymbol{y}_2) = (c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2) + (c_1\boldsymbol{y}_1 + c_2\boldsymbol{y}_2) \tag{6.5}$$

Since X is a **subspace** of V, $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 \in X$ and since Y is a subspace, $c_1\boldsymbol{y}_1 + c_2\boldsymbol{y}_2 \in Y$. By the definition of $Proj_{(X,Y)}$, (6.4), and (6.5) it follows that

$$Proj_{(X,Y)}(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = c_1Proj_{(X,Y)}(\boldsymbol{v}_1) + c_2Proj_{(X,Y)}(\boldsymbol{v}_2)$$

as we needed to show. □

Definition 6.2. Let V be a **vector space** , X and Y **subspaces** of V. Assume that $X \cap Y = \{\mathbf{0}_V\}$ and $X + Y = V$. We then say that V is the ***direct sum*** of X and Y and write $V = X \oplus Y$. We say $V = X \oplus Y$ is a ***direct sum decomposition of*** V.

Definition 6.3. Assume that $V = X \oplus Y$, a **direct sum** of the **subspaces** X and Y. The mapping $Proj_{(X,Y)}$ is called the ***projection map with respect to*** X ***and*** Y.

Remark 6.1. The ordering of X and Y makes a difference in the definition of $Proj_{(X,Y)}$ and, in fact, $Proj_{(X,Y)} \neq Proj_{(Y,X)}$.

We now begin to prove some properties of **linear transformations**.

Theorem 6.1.3. *Let $T : V \to W$ be a* **linear transformation**. *Then the following hold:*

1). $T(\mathbf{0}_V) = \mathbf{0}_W$; and

2) $T(\boldsymbol{u} - \boldsymbol{v}) = T(\boldsymbol{u}) - T(\boldsymbol{v})$.

Proof. 1) Since $\mathbf{0}_V + \mathbf{0}_V = \mathbf{0}_V$ we get

$$T(\mathbf{0}_V) = T(\mathbf{0}_V + \mathbf{0}_V) = T(\mathbf{0}_V) + T(\mathbf{0}_V).$$

Adding the opposite of $T(\mathbf{0}_V)$, $-T(\mathbf{0}_V)$, to both sides we get

$$\mathbf{0}_W = T(\mathbf{0}_V) + (-T(\mathbf{0}_V)) = [T(\mathbf{0}_V) + T(\mathbf{0}_V)] + (-T(\mathbf{0}_V)) =$$

$$T(\mathbf{0})_V) + [T(\mathbf{0}_V) + (-T(\mathbf{0}_V))] = T(\mathbf{0}_V) + \mathbf{0}_W = T(\mathbf{0}_V).$$

2) By **Theorem** (6.1.1) we have

$$T(\boldsymbol{u} - \boldsymbol{v}) = T((1)\boldsymbol{u} + (-1)\boldsymbol{v}) =$$

$$(1)T(\boldsymbol{u}) + (-1)T(\boldsymbol{v}) = T(\boldsymbol{u}) - T(\boldsymbol{v})$$

□

We next show that the **range** of a **linear transformation** T from a **vector space** V to a vector space W is a **subspace** of W.

Theorem 6.1.4. *Let $T : V \to W$ be a* **linear transformation** *of* **vector spaces***. Then $R(T)$, the* **range** *of T, is a* **subspace** *of W.*

Proof. Suppose that $\boldsymbol{w}_1, \boldsymbol{w}_2$ are in $R(T)$ and c_1, c_2 are scalars. We need to show that $c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 \in R(T)$. Now we have to remember what it means to be in $R(T)$. A vector $\boldsymbol{w}$ is in $R(T)$ if there is a vector $\boldsymbol{v} \in V$ such that $T(\boldsymbol{v}) = \boldsymbol{w}$. Since we are assuming that $\boldsymbol{w}_1, \boldsymbol{w}_2$ are in $R(T)$ there are vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ such that $T(\boldsymbol{v}_1) = \boldsymbol{w}_1, T(\boldsymbol{v}_2) = \boldsymbol{w}_2$. Since V is a **vector space** and $\boldsymbol{v}_1, \boldsymbol{v}_2$ are in V and c_1, c_2 are scalars, it follows that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$ is a vector in V. Now, $T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2$ by **Theorem** (6.1.1). So, $c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2$ is the **image** of the vector $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$ and hence in $R(T)$ as required. □

We shall show that a **linear transformation** is completely determined by its **images** on a **basis** . We proved this for a **linear transformation** T from $\mathbb{R}^n$ to $\mathbb{R}^m$ when the **basis** is the **standard basis** of $\mathbb{R}^n$. Here we will prove it in full generality, but before proceeding to the general statement and proof we do some examples.

Example 6.1.2. Suppose T is a **linear transformation** from $\mathbb{R}^3$ to $\mathbb{R}^4$ and we know the images of T on the **standard basis**, $(\boldsymbol{e}_1^2, \boldsymbol{e}_2^3, \boldsymbol{e}_3^3) = \left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right)$, for example, say

$$T(\boldsymbol{e}_1^3) = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, T(\boldsymbol{e}_2^3) = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, T(\boldsymbol{e}_3^3) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}.$$

Since $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\boldsymbol{e}_1^3 + y\boldsymbol{e}_2^4 + z\boldsymbol{e}_3^3$ we have $T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = T(x\boldsymbol{e}_1^3 + y\boldsymbol{e}_2^3 + z\boldsymbol{e}_3^3) = xT(\boldsymbol{e}_1^3) + yT(\boldsymbol{e}_2^3) + zT(\boldsymbol{e}_3^3) = x\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} + y\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} + z\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} x+y+z \\ -x \\ -y \\ -z \end{pmatrix}$. Thus, we know the **image** of any vector.

Example 6.1.3. Let $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix}\right)$. The **reduced echelon form** of the matrix $\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 2 \\ 1 & 1 & 2 \end{pmatrix}$ is the 3×3 **identity matrix** , I_3, from which we conclude by **Theorem** (3.4.15) that $\mathcal{B}$ is a **basis** for $\mathbb{R}^3$. Suppose T is a **linear transformation** from $\mathbb{R}^3 \to \mathbb{R}^2$ and we know the **image** of T on the vectors of $\mathcal{B}$. , For example assume

$$T\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, T\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, T\begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

We claim that we have enough information to find the **image** of T on any vector of $\mathbb{R}^3$. Now the **reduced echelon form** of the matrix $\begin{pmatrix} 1 & 1 & 3 & | & 1 & 0 & 0 \\ 1 & 2 & 2 & | & 0 & 1 & 0 \\ 1 & 1 & 2 & | & 0 & 0 & 1 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 & 0 & | & -2 & -1 & 4 \\ 0 & 1 & 0 & | & 0 & 1 & -1 \\ 0 & 0 & 1 & | & 1 & 0 & -1 \end{pmatrix}$ so that the **change of basis matrix** from $\mathcal{S}$ to $\mathcal{B}$,

$P_{\mathcal{S}\to\mathcal{B}} = \begin{pmatrix} -2 & -1 & 4 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix}$.

By the definition of the **change of basis matrix** it follows that the **coordinate vector** for $\boldsymbol{e}_1^3$, the first **standard basis vector** of $\mathbb{R}^3$, with respect to $\mathcal{B}$ is

$$[e_1]_{\mathcal{B}} = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

which means that

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

It then follows that

$$T(\boldsymbol{e}_1^3) = T((-2) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix}) = (-2)T \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + T \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix} = (-2) \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}.$$

In a similar fashion

$$T(\boldsymbol{e}_2^3) = T(- \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}) = -T \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + T \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = - \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and

$$T(\boldsymbol{e}_3^3) = T(4 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix}) = 4T \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - T \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - T \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix} =$$

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

It then follows that

$$T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = xT(\boldsymbol{e}_1^3) + yT(\boldsymbol{e}_2^3) + zT(\boldsymbol{e}_3^3) =$$

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

We now show that, in general, a **linear transformation** $T : V \to W$ is determined by its **images** on a **basis** of V.

Theorem 6.1.5. *Let V be a **vector space** with **basis** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ and let T be a **linear transformation** from V to a vector space W. Then T is completely determined by its **images** on $\mathcal{B}$. Moreover, the sequence $T(\mathcal{B}) = (T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \dots, T(\boldsymbol{v}_n))$ **spans** the **range** of T, $R(T)$.*

Proof. Let $\boldsymbol{v}$ be any vector in V and assume the **coordinate vector** of $\boldsymbol{v}$ with respect to $\mathcal{B}$, $[\boldsymbol{v}]_{\mathcal{B}}$ is $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$.

This means that $\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$. Then $T(\boldsymbol{v}) = T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \cdots + c_nT\boldsymbol{v}_n)$ which shows that the **image**, $T(\boldsymbol{v})$, of $\boldsymbol{v}$ is determined provided we know $T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n)$.

We get more: Note that $c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \cdots + c_nT(\boldsymbol{v}_n)$ is a **linear combination** of $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n))$ and therefore in $Span(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n))$ and consequently, the **range** of T is contained in $Span(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n))$.

On the other hand, each of the vectors $T(\boldsymbol{v}_i)$ is in $R(T)$. Since $R(T)$ is a **subspace** by **Theorem** (6.1.4) it then follows by **Theorem** (5.2.1) that $Span(T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \dots, T(\boldsymbol{v}_n))$ is contained in $R(T)$ and therefore we have equality. $\square$

What You Can Now Do

1. Determine if a given **function** is a **linear transformation**.

2. Given a **linear transformation** find the **domain and codomain** .

3. Given the **images** of a **linear transformation** $T : V \to W$ on the vectors of a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ and a vector $\boldsymbol{u}$ in V compute the **image** $T(\boldsymbol{u})$.

Method (How To Do It)

Method 6.1.1. Determine if a given **function** is a **linear transformation**.

Let $T : V \to W$ be a **function** from a **vector space** V to a vector space W. Typically the components of the **image** of T on a vector $\boldsymbol{v}$ will be given by functions of the components of $\boldsymbol{v}$. The function T will be a **linear transformation** if and only if each function is a linear expression without constant term. If there is a non-zero constant term we can demonstrate that T is not a **linear transformation** by showing the **image** of **zero vector** of V is not the **zero vector** of W which contradicts **Theorem** (6.1.3). If there are no non-zero constant terms but the expression is not a linear expression then we can show that the **scalar condition** does not hold.

Example 6.1.4. Let $T : \mathbb{R}_2[x] \to M_{2\times 2}(\mathbb{R})$ be given by

$$T(a + bx + cx^2) = \begin{pmatrix} a-2b & b-2c \\ a+b-c+2 & a-c \end{pmatrix}$$

This is not linear since the (2,1) entry is a linear expression of a, b, c but there is a non-zero constant term. Consequently, $T(\mathbf{0}_{\mathbb{R}_2[x]}) = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \neq \mathbf{0}_{M_{2\times 2}(\mathbb{R})}$.

Example 6.1.5. Let $T : \mathbb{R}^2 \to \mathbb{R}_2[x]$ be given by $T\begin{pmatrix} a \\ b \end{pmatrix} = (a^2 - b) + (2a + 3b)x + (3a + 4b)x^2$.

This function is not a **linear transformation** though there are no non-zero constant terms in any of the components because the first component is not a linear expression of a and b. We show the **scalar condition** fails:

$T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 + 2x + 3x^2$. However, $T\begin{pmatrix} 2 \\ 0 \end{pmatrix} = 4 + 4x + 6x^2 \neq 2(1 + 2x + 3x^2)$.

Example 6.1.6. Define $T : \mathbb{R}_1[x] \to \mathbb{R}_2[x]$ by $T(a + bx) = (2a - 3b) + (3ab)x + (-a + 2b)x^2$.

This function is not a **linear transformation** since the second component is not a linear expression of a and b. However, if we are to show that the **scalar condition** does not work we have to stay away from vectors with $a = 0$ or $b = 0$ since for such vectors the second component of the image becomes zero. So, we use $1 + x$.

$T(1+x) = -1+3x+x^2$. However, $T(2+2x) = -2+12x+2x^2 \neq 2(-1+3x+x^2)$.

Example 6.1.7. Let $T : M_{2\times 3}(\mathbb{R}) \to \mathbb{R}_2[x]$ be given by

$$T(\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}) = (a_{11} - a_{21}) + (a_{12} + 2a_{22})x + (-3a_{13} + 5a_{23})x^2$$

This function is a **linear transformation** since each components of the **image vector** is linear with zero constant term.

Method 6.1.2. Given a **linear transformation** find the **domain and codomain**.

This is a matter of figuring out the **vector space** where the function originates and the **vector space** where the **images** lie. This can usually be surmised from the formulas defining the **linear transformation**.

Example 6.1.8. Let T be the **linear transformation** given by

$$T(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}) = (2a_{11} - 3a_{12}) + (3a_{11} + 3a_{21})x + (3a_{12} - 3a_{21} + 4a_{22})x^2.$$

The vectors originate in $M_{2\times 2}(\mathbb{R})$ so this is the **domain** and the **images** all lie in $\mathbb{R}_2[x]$ and this is the **codomain**.

Example 6.1.9. Define T by

$$T(a_0 + a_1x + a_2x^2 + a_3x^3) = a_1 + 2a_2x + 3a_3x^2.$$

The **domain** is $\mathbb{R}_3[x]$ and the **codomain** is $\mathbb{R}_2[x]$.

Example 6.1.10. For a real valued function f of a real variable define $T(f) = f(1)$.

The **domain** is $F(\mathbb{R}, \mathbb{R})$ the **vector space** of all real valued functions of a real variable and the **codomain** is the one dimensional space $\mathbb{R}$.

Example 6.1.11. Let T be given by

$$T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 & x_1+2x_2 & x_1+3x_3 \\ x_1+2x_2 & 4x_2 & 2x_2+3x_3 \\ x_1+3x_3 & 2x_2+3x_3 & 6x_3 \end{pmatrix}$$

The **domain** is $\mathbb{R}^3$ and the **codomain** is $M_{3\times 3}(\mathbb{R})$.

Method 6.1.3. Given the **images** of a **linear transformation** $T : V \to W$ on the vectors of a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ of V and a vector $\boldsymbol{u}$ in V compute the **image** $T(\boldsymbol{u})$.

Suppose $T : V \to W$ and $T(\boldsymbol{v}_i) = \boldsymbol{w}_i$ are given. Use **Method** (5.4.1) to find the **coordinate vector** of $\boldsymbol{u}$ with respect to $\mathcal{B}$, $[\boldsymbol{u}]_{\mathcal{B}}$. Say $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$. Then $\boldsymbol{u} = u_1\boldsymbol{v}_1 + u_2\boldsymbol{v}_2 + \cdots + u_n\boldsymbol{v}_n$.

Since T is a **linear transformation** we have

$$T(\boldsymbol{u}) = T(u_1\boldsymbol{v}_1 + u_2\boldsymbol{v}_2 + \cdots + u_n\boldsymbol{v}_n) =$$

$$u_1T(\boldsymbol{v}_1) + u_2T(\boldsymbol{v}_2) + \cdots + u_nT(\boldsymbol{v}_n) = u_1\boldsymbol{w}_1 + u_2\boldsymbol{w}_2 + \cdots + u_n\boldsymbol{w}_n.$$

Example 6.1.12. Assume that $T : \mathbb{R}^3 \to M_{2\times 2}(\mathbb{R})$ is a **linear transformation** and

$$T\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & -6 \end{pmatrix},$$

$$T\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 5 & -9 \end{pmatrix},$$

$$T\begin{pmatrix}3\\2\\3\end{pmatrix} = \begin{pmatrix}3 & 5\\8 & -16\end{pmatrix}.$$

a) Verify that $\mathcal{B} = \left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\2\\2\end{pmatrix}, \begin{pmatrix}3\\2\\3\end{pmatrix}\right)$ is a **basis** for $\mathbb{R}^3$; and

b) Find the **image** of $\boldsymbol{u} = \begin{pmatrix}1\\3\\6\end{pmatrix}$.

We verify that $\mathcal{B}$ is a **basis** and find $[\boldsymbol{u}]_{\mathcal{B}}$, the **coordinate vector** of $\boldsymbol{u}$ with respect to $\mathcal{B}$, simultaneously by applying **Gaussian elimination** to the matrix $\begin{pmatrix}1 & 1 & 3 & | & 1\\1 & 2 & 2 & | & 3\\1 & 2 & 3 & | & 6\end{pmatrix}$.

The **reduced echelon form** of this matrix is $\begin{pmatrix}1 & 0 & 0 & | & -13\\0 & 1 & 0 & | & 5\\0 & 0 & 1 & | & 3\end{pmatrix}$.

Consequently, $\mathcal{B}$ is a basis and $[\boldsymbol{u}]_{\mathcal{B}} = \begin{pmatrix}-13\\5\\3\end{pmatrix}$.

Thus, $T(\boldsymbol{u}) = (-13)\begin{pmatrix}1 & 2\\3 & -6\end{pmatrix} + 5\begin{pmatrix}1 & 3\\5 & -9\end{pmatrix} + 3\begin{pmatrix}3 & 5\\8 & -16\end{pmatrix} = \begin{pmatrix}1 & 4\\10 & -15\end{pmatrix}$.

Example 6.1.13. Assume that $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}_4[x]$ is a **linear transformation** and that

$$T\begin{pmatrix}1 & 2\\2 & 3\end{pmatrix} = 4 - 4x + x^2 + 4x^3 + 3x^4,$$

$$T\begin{pmatrix}1 & 4\\3 & 2\end{pmatrix} = 3 - 7x + x^2 + 7x^3 + 2x^4,$$

$$T\begin{pmatrix}2 & 3\\3 & 7\end{pmatrix} = 9 - 6x + 2x^2 + 6x^3 + 7x^4,$$

$$T\begin{pmatrix}2 & 5\\4 & 5\end{pmatrix} = 7 - 9x + 2x^2 + 9x^3 + 5x^4.$$

a) Verify that

$$\mathcal{B} = \left(\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ 3 & 7 \end{pmatrix}, \begin{pmatrix} 2 & 5 \\ 4 & 5 \end{pmatrix} \right)$$

is a **basis** for $M_{2\times 2}(\mathbb{R})$; and

b) Find the T- **images** of the **standard basis vectors** of $M_{2\times 2}(\mathbb{R})$.

We verify that $\mathcal{B}$ is a **basis** and find the **coordinate vectors** of the **standard basis vectors** of $M_{2\times 2}(\mathbb{R})$ all at once by applying **Gaussian elimination** to the matrix obtained by first writing the vectors of $\mathcal{B}$ as columns followed by the vectors of the standard basis written as columns. The matrix is

$$\begin{pmatrix} 1 & 1 & 2 & 2 & | & 1 & 0 & 0 & 0 \\ 2 & 3 & 3 & 4 & | & 0 & 1 & 0 & 0 \\ 2 & 4 & 3 & 5 & | & 0 & 0 & 1 & 0 \\ 3 & 2 & 7 & 5 & | & 0 & 0 & 0 & 1 \end{pmatrix}$$

This matrix has **reduced echelon form**:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & 2 & 4 & -3 & -1 \\ 0 & 1 & 0 & 0 & | & -5 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & | & -3 & -1 & 1 & 1 \\ 0 & 0 & 0 & 1 & | & 5 & -1 & 0 & -1 \end{pmatrix}$$

This verifies that $\mathcal{B}$ is a **basis** and that

$$[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}]_{\mathcal{B}} = \begin{pmatrix} 2 \\ -5 \\ -3 \\ 5 \end{pmatrix},$$

$$[\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}]_{\mathcal{B}} = \begin{pmatrix} 4 \\ 0 \\ -1 \\ -1 \end{pmatrix},$$

$$[\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}]_{\mathcal{B}} = \begin{pmatrix} -3 \\ 1 \\ 1 \\ 0 \end{pmatrix},$$

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

Then

$$T(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}) = 2(4-4x+x^2+4x^3+3x^4)+(-5)(3-7x+x^2+7x^3+2x^4)+$$

$$(-3)(9-6x+2x^2+6x^3+7x^4)+5(7-9x+2x^2+9x^3+5x^4) = 1+x^2,$$

$$T(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}) = 4(4-4x+x^2+4x^3+3x^4)+0(3-7x+x^2+7x^3+2x^4)+$$

$$(-1)(9-6x+2x^2+6x^3+7x^4)+(-1)(7-9x+2x^2+9x^3+5x^4) = -x+x^3,$$

$$T(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}) = (-3)(4-4x+x^2+4x^3+3x^4)+1(3-7x+x^2+7x^3+2x^4)+$$

$$1(9-6x+2x^2+6x^3+7x^4)+0(7-9x+2x^2+9x^3+5x^4) = -x+x^3,$$

$$T(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}) = (-1)(4-4x+x^2+4x^3+3x^4)+1(3-7x+x^2+x^3+2x^4)+$$

$$1(9-6x+2x^2+6x^3+7x^4)+(-1)(7-9x+2x^2+9x^3+5x^4) = -1+x^4.$$

Exercises

In 1 - 7 determine if the function is a **linear transformation**. If it is non-linear demonstrate that this is so. See **Method** (6.1.1).

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

2. $T(a+bx+cx^2) = \begin{pmatrix} ab \\ a-b \\ b-c \end{pmatrix}$

3. $T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bcx.$

4. $T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \end{pmatrix}.$

5. $T(a+bx+cx^2) = 1 + \frac{a}{2}x^2 + \frac{b}{6}x^3 + \frac{c}{12}x^4.$

6. $T\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = 1 + (a-b)x + (b-c)x^2 + (d-c)x^3.$

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

In exercises 8-14 find the **domain and codomain** of the given **linear transformation**. See **Method** (6.1.2).

8. The transformation of exercise 1.

9. The transformation of exercise 2.

10. Transformation of exercise 3.

11. The transformation of exercise 4.

12. The transformation of exercise 5.

13. The transformation of exercise 6.

14. Transformation of exercise 7.

In exercises 15 and 16: a) Verify the given sequence $\mathcal{B}$ is a **basis** ; and b) compute the **images** of the standard basis vectors with respect to the given **linear transformation**. See **Method** (6.1.3).

15. $\boldsymbol{f}_1 = 1 + 2x - 2x^2, \boldsymbol{f}_2 = 2 + 3x - 4x^2, \boldsymbol{f}_3 = 1 + 4x - x^2, \mathcal{B} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3).$

$$T(\boldsymbol{f}_1) = \begin{pmatrix} 1 & 1 \\ 3 & 5 \end{pmatrix}, T(\boldsymbol{f}_2) = \begin{pmatrix} 2 & 1 \\ 5 & 8 \end{pmatrix}, T(\boldsymbol{f}_3) = \begin{pmatrix} 1 & 4 \\ 5 & 10 \end{pmatrix}.$$

16. $\mathcal{B} = (\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4)$ where

$$\boldsymbol{m}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \boldsymbol{m}_2 = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}, \boldsymbol{m}_3 = \begin{pmatrix} 1 & 3 \\ 2 & 3 \end{pmatrix}, \boldsymbol{m}_4 = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$$

$$T(\boldsymbol{m}_1) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, T(\boldsymbol{m}_2) = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 2 \end{pmatrix}, T(\boldsymbol{m}_3) = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 3 \end{pmatrix}, T(\boldsymbol{m}_4) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 4 \end{pmatrix}.$$

In exercises 17 - 20 answer true or false and give an explanation.

17. If $T : \mathbb{R}^4 \to M_{2\times 2}(\mathbb{R})$ is a **linear transformation** and $T(\boldsymbol{e}_j) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ for $j = 1, 2, 3, 4$ then $T(\boldsymbol{x}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ for every vector $\boldsymbol{x} \in \mathbb{R}^4$.

18. If $T : V \to W$ is a **linear transformation** and $T(\boldsymbol{x}) = T(\boldsymbol{y})$ then $T(\boldsymbol{x} - \boldsymbol{y}) = \mathbf{0}_W$.

19. If $T : V \to W$ is a **linear transformation** and $\boldsymbol{w} \notin R(T)$ then also $2\boldsymbol{w} \notin R(T)$.

20. If $T : \mathbb{R}[x]_2 \to \mathbb{R}^3$ is a **linear transformation** and $R(T) = \mathbb{R}^3$ then $T(1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, T(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, T(x^2) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

Challenge Exercises (Problems)

1. Let $S : U \to V$ and $T : V \to W$ be **linear transformations**. Define the composition of these transformations, $T \circ S : U \to W$ by $T \circ S(\boldsymbol{u}) = T(S(\boldsymbol{u}))$.

Prove that $T \circ S$ is a l **linear transformation**.

2. Assume $dim(V) = 4, dim(W) = 3$ and $T : V \to W$ is a **linear transformation**. Prove that there exists a non-zero vector $\boldsymbol{v} \in V$ such that $T(\boldsymbol{v}) = \mathbf{0}_W$.

3. Let $T : V \to W$ be a **linear transformation**. Set $K(T) := \{\boldsymbol{v} \in V : T(\boldsymbol{v}) = \mathbf{0}_W\}$. Prove that $K(T)$ is a **subspace** of V.

4. We continue with the hypotheses of CE 3.

a) Let $\boldsymbol{x}, \boldsymbol{y} \in V$ satisfy $T(\boldsymbol{x}) = T(\boldsymbol{y})$. Prove that $\boldsymbol{x} - \boldsymbol{y} \in K(T)$.

b) Let $\boldsymbol{x} \in V, \boldsymbol{u} \in K(T)$. Prove that $T(\boldsymbol{x} + \boldsymbol{u}) = T(\boldsymbol{x})$.

c) Assume $T(\boldsymbol{x}) = \boldsymbol{w}$. Prove that $\{\boldsymbol{v} \in V : T(\boldsymbol{v}) = \boldsymbol{w}\} = \boldsymbol{x} + K(T)$.

Quiz Solutions

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

The **domain** is $\mathbb{R}^3$ and the **codomain** is $M_{2\times 2}(\mathbb{R})$.

Not right, see **concepts related to the notion of a function**.

2. $T(a + bx + cx^2) = ax + \frac{b}{2}x^2 + \frac{c}{3}x^3$.

The **domain** is $\mathbb{R}_2[x]$ and the **codomain** is $\mathbb{R}_3[x]$.
Not right, see **concepts related to the notion of a function**.

3. $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x+3y \\ 3x+4y-1 \\ x+y \end{pmatrix}$

Because the second component is not a linear expression **without** constant term it is not a matrix transformation. Specifically, $T(\mathbf{0}_2) \neq \mathbf{0}_3$.

$T(\mathbf{0}_2) = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}$

Not right, see **Theorem** (3.1.4).

4. $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ xy \end{pmatrix}$

Now the third component is not linear expression of x and y. As a consequence the scalar condition will not be satisfied:

$T\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. However, $T2\begin{pmatrix} 1 \\ 1 \end{pmatrix} = T\begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} \neq 2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Alternatively, we can show that the additive property of a matrix transformation does not hold:

$T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

However, $T(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}) = T\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = T\begin{pmatrix} 1 \\ 0 \end{pmatrix} + T\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Not right, see **definition of a matrix transformation**.

5. Find the **standard matrix** of the following transformation:

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+2y+3z \\ 2x+y-4z \\ 3x+7y-z \end{pmatrix} =$$

$$\begin{pmatrix} x \\ 2x \\ 3x \end{pmatrix} + \begin{pmatrix} 2y \\ y \\ 7y \end{pmatrix} + \begin{pmatrix} 3z \\ -4z \\ -z \end{pmatrix} =$$

$$x\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + y\begin{pmatrix} 2 \\ 1 \\ 7 \end{pmatrix} + z\begin{pmatrix} 3 \\ -4 \\ -1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 3 & 7 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

The **standard matrix** is $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 3 & 7 & -1 \end{pmatrix}$.

Not right, see **Method** (3.1.2).

6. We do a) and b) together using **Method** (5.2.3) and **Method** (5.2.6). Thus, with each vector $\boldsymbol{m} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ we associate a **4-vector**, $\begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{pmatrix}$. We then make a matrix with the the associated with $\mathcal{B}$ as the first four columns followed by the vectors associated to the **standard basis vectors** for $M_{2\times 2}(\mathbb{R})$, namely, the **identity matrix**, I_4. We then use **Gaussian elimination** to obtain the **reduced echelon form**. If $\mathcal{B}$ is a **basis** on the left hand side of the matrix we should obtain I_4. Then on the right hand side will be the respective **coordinate vectors**. The matrix is

$$A = \left(\begin{array}{cccc|cccc} 1 & 1 & 3 & 3 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 2 & 2 & 3 & 0 & 0 & 1 & 0 \\ 1 & 3 & 1 & 4 & 0 & 0 & 0 & 1 \end{array}\right)$$

The **reduced echelon form** of this matrix is

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 3 & 4 & -7 & 1 \\ 0 & 1 & 0 & 0 & -2 & -1 & 4 & -1 \\ 0 & 0 & 1 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & 1 & 1 & 0 & -2 & 1 \end{array}\right)$$

This verifies that $\mathcal{B}$ is a **basis** for $M_{2\times 2}(\mathbb{R})$. Moreover,

$$[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}]_{\mathcal{B}} = \begin{pmatrix} 3 \\ -2 \\ -1 \\ 1 \end{pmatrix}$$

$$[\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}]_{\mathcal{B}} = \begin{pmatrix} 4 \\ -1 \\ -1 \\ 0 \end{pmatrix}$$

$$[\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}]_{\mathcal{B}} = \begin{pmatrix} -7 \\ 4 \\ 3 \\ -2 \end{pmatrix}$$

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

Not right, see **Method** (5.2.3) and **Method** (5.2.6).

.

6.2. Range and Kernel of a Linear Transformation

In this section we define and study the kernel of a linear transformation. The kernel consists of all vectors in the domain which get sent to the zero vector. We also study the range of a linear transformation.

The notion of a linear transformation on abstract vector spaces is introduced and we establish a number of important properties.

Am I Ready for This Material

Familiarity with the following concepts is essential to success in mastering the material of this section:

null space of a matrix

span of a sequence of vectors

linearly independent sequence of vectors

basis of a vector space

dimension of a vector space

column space of a matrix

rank of a matrix

linear transformation

range of a function

Quiz

In 1 and 2 for the given matrices find a) a **basis** for the **column space**, b) a **basis** for the **row space**, c) a **basis** for the **null space**. Determine the **rank** and **nullity** of the matrix. Decide if the sequence of columns **span** the appropriate $\mathbb{R}^n$ and if the sequence of columns is **linearly independent**.

1. $\begin{pmatrix} 1 & 2 & 1 & 1 & 0 \\ 1 & 3 & 0 & 2 & 1 \\ 2 & 5 & 1 & 4 & 1 \\ 3 & 5 & 4 & 2 & -1 \end{pmatrix}$

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

Quiz Solutions

New Concepts

Several very important concepts are introduced in this section all of which are central to the study of **linear transformations**.

The last concept will make formal the intuitive sense that all **vector spaces** of **dimension** n are "alike." These concepts are:

kernel of a linear transformation

nullity of a linear transformation

rank of a linear transformation

surjective or onto transformation

injective or one-to-one transformation

bijective transformation

one-to-one correspondence between two sets

isomormphism of vector spaces

isomorphic vector spaces

Theory (Why It Works)

Let A be the matrix $\begin{pmatrix} 1 & -1 & 2 & -2 \\ 2 & 3 & -1 & -4 \\ 4 & 1 & 3 & -8 \\ 1 & 4 & -3 & -2 \end{pmatrix}$ and $T = T_A : \mathbb{R}^4 \to \mathbb{R}^4$ the **linear transformation** defined by $T(\boldsymbol{x}) = A\boldsymbol{x}$. We wish to determine the collection of vectors $\boldsymbol{x}$ such that $T(\boldsymbol{x}) = \mathbf{0}_4$. Of course, this is just the **null space** of the matrix A. We will be interested in the same question for an arbitrary **linear transformation** between **vector spaces** and this motivates the following definition:

Definition 6.4. Let $T : V \to W$ be a **linear transformation**. The ***kernel*** of T, denoted by $Ker(T)$, consists of all vectors in V which go to the **zero vector** of W : $Ker(T) := \{\boldsymbol{v} \in V | T(\boldsymbol{v}) = \mathbf{0}_W\}$.

Recall, we defined the **range** of T, denoted by $R(T)$, to be the set of all **images** of T :

$$R(T) = \{T(\boldsymbol{v}) | \boldsymbol{v} \in V\}.$$

When $T : V \to W$ is a **linear transformation** we proved in **Theorem** (6.1.4) that $R(T)$ is a **subspace** of W. In our next theorem we prove that $Ker(T)$ is a **subspace** of V.

Theorem 6.2.1. *Let $T : V \to W$ be a* **linear transformation***. Then $Ker(T)$ is a* **subspace** *of V.*

Proof. Assume $\boldsymbol{v}_1, \boldsymbol{v}_2$ are in $Ker(T)$ and c_1, c_2 are scalars. By **Theorem** (6.1.1) we need to show that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$ is in $Ker(T)$. It is important to remind ourselves what this means: a vector $\boldsymbol{v} \in Ker(T)$ if and only if $T(\boldsymbol{v}) = \mathbf{0}_W$. Since we are assuming that $\boldsymbol{v}_1, \boldsymbol{v}_2$ are in $Ker(T)$ this means that $T(\boldsymbol{v}_1) = T(\boldsymbol{v}_2) = \mathbf{0}_W$. Let's apply T to $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$:

$$T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) =$$

$$c_1\mathbf{0}_W + c_2\mathbf{0}_W =$$

$$\mathbf{0}_W + \mathbf{0}_W = \mathbf{0}_W.$$

So, $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$ is in $Ker(T)$ as required. □

Example 6.2.1. 1) Let $D : \mathbb{R}_3[x] \to \mathbb{R}_2[x]$ be the derivative. Then **Ker(D)** $= \mathbb{R}$, $R(D) = \mathbb{R}_2[x]$.

2) Let D^2 be the map from the space of twice differentible functions to $F[\mathbb{R}]$ given by $D^2(f) = \frac{d^2f}{dx^2}$. What is the **kernel** of $D^2 + D$.

These are the functions which satisfy the second order differenial equation

$$\frac{d^2f(x)}{dx^2} + f(x) = 0.$$

In a course in ordinary differential equations one proves this is a two **dimensional** vector space with basis $(sin\ x, cos\ x)$.

3) Let $A = \begin{pmatrix} 1 & -2 & 1 \\ -1 & 3 & 2 \\ -1 & 1 & 0 \end{pmatrix}$ and let $T(\begin{pmatrix} x \\ y \\ z \end{pmatrix}) = A\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Then $Ker(T) = null(A)$ and $R(T) = col(A)$.

Since the **range** and the **kernel** of a **linear transformation** are **subspaces** each has a **dimension**. For further reference we give names to these dimensions:

Definition 6.5. Let V and W be **finite dimensional vector spaces** and $T : V \to W$ a **linear transformation**. We will refer to the **dimension** of the **range** of T as the ***rank of*** T and denote this by $Rank(T)$. Thus, $Rank(T) = dim(R(T))$.

Definition 6.6. Let V and W be **finite dimensional vector space** and $T : V \to W$ a **linear transformation**. We will refer to the **dimension** of the **kernel** of T as the ***nullity of*** T. We denote this by $Nullity(T)$. Thus, $Nullity(T) = dim(Ker(T))$.

Our next result identifies the **kernel** and **range** of a **matrix transformation** as familiar objects:

Theorem 6.2.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **matrix transformation** with **standard matrix** A. Then*

1) $R(T) = col(A)$.
2) $Ker(T) = null(A)$.
3) $Rank(T) = rank(A)$.
4) $Nullity(T) = nullity(A)$.

Proof. 1) Suppose $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$. Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ be the columns of A. Then $T(\boldsymbol{x}) = A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$, a **linear combination** of the columns and hence an element of $Span(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n) = Col(A)$. Since $\boldsymbol{x}$ is arbitrary we get every such **linear combinations**.

2), 3) and 4) follow from the definitions. $\square$

Example 6.2.2. Find the **kernel**, **range**, **nullity** and **rank** of the **matrix transformation** with **standard matrix** $A = \begin{pmatrix} 1 & -1 & 2 & -2 \\ 2 & 3 & -1 & -4 \\ 4 & 1 & 3 & -8 \\ 1 & 4 & -3 & -2 \end{pmatrix}$.

By obtaining the **reduced echelon form** we can find a **basis** for the column space and express the **range** as a span of a basis. We can also get a **basis** for the **kernel**. The **reduced echelon form** is $R = \begin{pmatrix} 1 & 0 & 1 & -2 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

Thus, the first two columns are **pivot columns** and therefore the first two columns of A is a **basis** for the **column space** of A and, consequently, $R(T) = Span\left(\begin{pmatrix}1\\2\\4\\1\end{pmatrix}, \begin{pmatrix}-1\\3\\1\\4\end{pmatrix}\right)$ and $Rank(T) = 2$.

The **homogeneous linear system** which has **coefficient matrix** R is

$$\begin{array}{ccccccccc} x_1 & & & + & x_3 & - & 2x_4 & = & 0 \\ & & x_2 & - & x_3 & & & = & 0 \end{array}$$

There are two **leading variables** (x_1, x_2) and two **free variables** (x_3, x_4). The general solution is given parametrically by

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

Thus

$$Ker(T) = Span\left(\begin{pmatrix}-1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}2\\0\\0\\1\end{pmatrix}\right).$$

It follows that $Nullity(T) = 2$.

The next result is an extension of **Theorem** (5.5.4) which we proved for **matrix transformations** to an arbitrary **linear transformation** on a **finite dimensional vector space** and is among the most important theorems in all of (elementary) linear algebra. We refer to it as the **"Rank-Nullity Theorem for Linear Transformations".**

Theorem 6.2.3. *Let V be an* **n-dimensional vector space** *and $T : V \to W$ be a* **linear transformation**. *Then $n = dim(V) = Rank(T) + Nullity(T)$.*

Proof. Let $k = Nullity(T)$. Choose a basis $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ for $Ker(T)$. Extend this to a basis $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ for V. We claim two things:

1) $(T(\boldsymbol{v}_{k+1}), \ldots, T(\boldsymbol{v}_n))$ is **linearly independent**; and

2) $(T(\boldsymbol{v}_{k+1}), \dots, T(\boldsymbol{v}_n))$ **spans** $R(T)$.

If both of these are true then the result follows since in this case $(T(\boldsymbol{v}_{k+1}), \dots, T(\boldsymbol{v}_n))$ is a **basis** for $R(T)$ and we will have $Rank(T) = n - k$ as required. So let us prove the two claims.

1) The first thing we prove is that

$$Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k) \cap Span(\boldsymbol{v}_{k+1}, \boldsymbol{v}_{k+2}, \dots, \boldsymbol{v}_n) = \{\boldsymbol{0}_V\}.$$

This will follow from the fact that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ is a **basis** and therefore **linearly independent**. To see this, suppose

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots c_k\boldsymbol{v}_k = c_{k+1}\boldsymbol{v}_{k+1} + \dots + c_n\boldsymbol{v}_n.$$

Then $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots c_k\boldsymbol{v}_k - c_{k+1}\boldsymbol{v}_{k+1} - \dots - c_n\boldsymbol{v}_n = \boldsymbol{0}_V$. Since $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ is **linearly independent** we must have $c_1 = c_2 = \dots = c_n = 0$ as claimed.

Suppose now that

$$c_{k+1}T(\boldsymbol{v}_{k+1}) + \dots + c_nT(\boldsymbol{v}_n) = \boldsymbol{0}_W.$$

Since this is the image of $\boldsymbol{u} = c_{k+1}\boldsymbol{v}_{k+1} + \dots + c_n\boldsymbol{v}_n$ the vector $\boldsymbol{u}$ is in $Ker(T)$. But then $c_{k+1}\boldsymbol{v}_{k+1} + \dots + c_n\boldsymbol{v}_n$ is in $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ and so in the intersection $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k) \cap Span(\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n)$ which we just proved is the **zero vector** of V.

Therefore $c_{k+1}\boldsymbol{v}_{k+1} + \dots + c_n\boldsymbol{v}_n = \boldsymbol{0}_V$ and since the sequence $(\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n)$ is **linearly independent** (it is a subsequence of a linearly independent sequence), it follows that $c_{k+1} = c_{k+2} = \dots = c_n = 0$. This proves that the sequence $(T(\boldsymbol{v}_{k+1}), \dots, T(\boldsymbol{v}_n))$ is **linearly independent** as claimed.

2) Since every vector in V is a linear combination of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ it follows that the typical element of the $R(T)$ is $T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_n\boldsymbol{v}_n) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \dots + c_kT(\boldsymbol{v}_k) + c_{k+1}T(\boldsymbol{v}_{k+1}) + \dots c_nT(\boldsymbol{v}_n)$. However, since $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k \in Ker(T)$ this is equal to $c_{k+1}T(\boldsymbol{v}_{k+1}) + \dots c_nT(\boldsymbol{v}_n)$ which is just an element of $Span(T(\boldsymbol{v}_{k+1}), \dots T(\boldsymbol{v}_n))$ as claimed. This completes the proof. $\square$

Before proceeding to some further results we introduce (review) the concepts of onto and one-to-one functions.

Definition 6.7. Let X and Y be sets. A function (map, transformation) $T : X \to Y$ is said to be ***onto or surjective*** if every element of Y is an **image** of some element in x. More succinctly, this means that $Y = R(T)$, that is, the **range** of T is equal to the **codomain** of T.

Definition 6.8. Let X and Y be sets. A function (map, transformation) $T : X \to Y$ is ***one-to-one or injective*** if, whenever y is in the **range** of $T, y \in R(T)$, then there is a **unique** $x \in X$ with $T(x) = y$. Thus, if $x, x' \in X$ and $T(x) = T(x')$ then $x = x'$. Still another way to put this is: if $x \neq x'$ then $T(x) \neq T(x')$.

Definition 6.9. Let $T : X \to Y$ be a function (map, transformation). T is a ***bijection*** or a ***one-to-one correspondence*** if it is both **surjective** and **injective**.

Example 6.2.3. 1. The map $exp : \mathbb{R} \to \mathbb{R}^+$ given by $exp(x) = e^x$ is a correspondence (by $\mathbb{R}^+$ we mean the positive real numbers).

2. The map $\sqrt{\ } : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ given by $x \to \sqrt{x}$ is injective (one-to-one) but not surjective (onto).

3. The function $x \to x^2$ is surjective (onto) $\mathbb{R}^+ \cup \{0\}$ but is not injective (one-to-one.)

4. The map $T : \mathbb{R} \to \mathbb{R}$ given by $T(x) = x^2 + 3x + 7$ is neither injective (one-to-one) nor surjective (onto).

5. The map $T : \mathbb{R} \to \mathbb{R}$ given by $T(x) = 2x + 3$ is a correspondence.

There is a beautiful criterion for a **linear transformation** to be **injective** which we establish in our next theorem.

Theorem 6.2.4. *Let $T : V \to W$ be a* **linear transformation**. *Then T is* **injective or one-to-one** *if and only if $Ker(T) = \{\mathbf{0}_V\}$.*

Proof. Suppose T is **one-to-one**. Then there is at most one vector $\boldsymbol{v} \in V$ such that $T(\boldsymbol{v}) = \mathbf{0}_W$. Since $\mathbf{0}_V$ maps to $\mathbf{0}_W$ it follows that $Ker(T) = \{\mathbf{0}_V\}$.

On the other hand suppose $Ker(T) = \{\mathbf{0}_V\}$, $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, and $T(\boldsymbol{v}_1) = T(\boldsymbol{v}_2)$. We need to prove that $\boldsymbol{v}_1 = \boldsymbol{v}_2$. Since $T(\boldsymbol{v}_1) = T(\boldsymbol{v}_2)$ it follows that $T(\boldsymbol{v}_1) - T(\boldsymbol{v}_2) = \mathbf{0}_W$. But $T(\boldsymbol{v}_1) - T(\boldsymbol{v}_2) = T(\boldsymbol{v}_1 - \boldsymbol{v}_2)$ and consequently $\boldsymbol{v}_1 - \boldsymbol{v}_2 \in Ker(T)$. But then $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \mathbf{0}_V$, whence $\boldsymbol{v}_1 = \boldsymbol{v}_2$ as desired. $\square$

Example 6.2.4. (1) Let $E : \mathbb{R}_2[x] \to \mathbb{R}^3$ be the transformation given by $E(f) = \begin{pmatrix} f(1) \\ f(2) \\ f(3) \end{pmatrix}$. This **linear transformation** is **one-to-one**.

(2) Consider the transformation $T : \mathbb{R}_2[x] \to \mathbb{R}^2$ given by $T(f) = \begin{pmatrix} f(1) \\ f(2) \end{pmatrix}$. Now, $Ker(T) = Span((x-1)(x-2))$.

(3) Consider $R : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}^2$ given by $R(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}) = \begin{pmatrix} a_{11} + a_{12} \\ a_{21} + a_{22} \end{pmatrix}$.

$$\textbf{Ker(R)} = Span\left(\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \right).$$

(4) Now map $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ to $\begin{pmatrix} a_{11} + a_{12} \\ a_{21} + a_{22} \\ a_{11} + a_{21} \\ a_{12} - a_{22} \end{pmatrix}$. This **linear transformation** **one-to-one**.

In our next theorem we demonstrate that a **linear transformation** which is **injective** maps a **linearly independent** sequence to a linearly independent sequence.

Theorem 6.2.5. *Assume $T : V \to W$ is a* **linear transformation**.

1) Assume that T is **injective**. *If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is a* **linearly independent** *sequence in V then $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_k))$ is* **linearly independent**.

2) Let $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ be a **basis** *of V. If $(T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \dots, T(\boldsymbol{v}_n))$ is* **linearly independent** *then T is* **injective**.

Proof. 1) Consider a **dependence relation** on $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_k))$: Suppose for the scalars $c_1, c_2, \dots, c_k$ that $c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \dots c_kT(\boldsymbol{v}_k) = \mathbf{0}_W$. We need to show that $c_1 = c_2 = \dots = c_k = 0$. Because T is a **linear transformation** ,

$$\mathbf{0}_w = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \dots + c_kT(\boldsymbol{v}_k) = T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_k\boldsymbol{v}_k)$$

This implies that the vector $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_k\boldsymbol{v}_k$ is in $Ker(T)$. However, by hypothesis, Ker(T) $= \{\mathbf{0}_V\}$. Therefore, $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots + c_k\boldsymbol{v}_k = \mathbf{0}_V$. But we are also assuming that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is **linearly independent**. Consequently, $c_1 = c_2 = \dots = c_k = 0$ as required.

2) Let $\boldsymbol{u} \in Ker(T)$. We must show that $\boldsymbol{u} = \mathbf{0}_V$. Since $\mathcal{B}$ is a **basis** there are scalar $c_1, c_2, \dots, c_n$ such that

$$\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n.$$

Since $\boldsymbol{u} \in Ker(T)$, $T(\boldsymbol{u}) = \boldsymbol{0}_W$ and by our properties of linear transformations this implies that

$$T(\boldsymbol{u}) = T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n) =$$

$$c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \cdots + c_nT(\boldsymbol{v}_n) = \boldsymbol{0}_W.$$

We are assuming that $(T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \dots, T(\boldsymbol{v}_n))$ is **linearly independent** and consequently $c_1 = c_2 = \cdots = c_n = 0$. Therefore $\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n = \boldsymbol{0}_V$ as required. □

In some of the examples above you may have noticed that when $T : V \to W$ is a **linear transformation** and $dim(V) = dim(W)$ then T is **one-to-one** appears to imply T is **onto** and vice versa. This is, in fact, true and the subject of our next theorem. We refer to this as the **"Half Is Good Enough Theorem For Linear Transformations".**

Theorem 6.2.6. *Let V and W be* **n-dimensional** *vector spaces and $T : V \to W$ be a* **linear transformation**. *Then T is* **one-to-one** *if and only if T is* **onto**.

Proof. Suppose T is **one-to-one**. Let $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ be a **basis** for V. By **Theorem** (6.2.5), $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n))$ is **linearly independent** in W. By **Theorem** (5.3.8), $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n))$ **spans** W. Then $W = Span((T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n)) = R(T)$ which proves that T is **onto**.

Converely, assume now that T is **onto**. Then $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n))$ **spans** W. By **Theorem** (5.3.8), the sequence, $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n))$ is **linearly independent** and by **Theorem** (6.2.5) T is **one-to-one**. □

We give a special name to **linear transformations** that are **bijective** and to the pairs of **vector spaces** that are connected by such transformations.

Definition 6.10. A **linear transformation** $T : V \to W$ which is **bijective** (**one-to-one** and **onto**) is said to be an ***isomorphism of vector spaces*** and the **vector spaces** V and W are ***isomorphic***.

The next theorem validates the intuitive sense that vector spaces like $\mathbb{R}^4$, $M_{2\times 2}(\mathbb{R}), \mathbb{R}_3[x]$ are alike (and the tendency to treat them as if they are identical).

Theorem 6.2.7. *Two* **finite dimensional vector spaces** V *and* W *are* **isomorphic** *if and only if* $dim(V) = dim(W)$.

Proof. Let T be an $m-$dimensional vector space, W and n-dimensional vector space, and assume $T : V \to W$ is an **isomorphism**. Since T is an isomorphism, T is **one-to-one** and **onto**. Since T is one-to-one we can conclude that $Ker(T)$ = $\{\mathbf{0}_V\}$, has dimension zero, and $Nullity(T)$ = 0. Since T is onto we can conclude that $Rank(T)$ = n. It now follows from **Theorem** (6.2.3) that $m = dim(V) = Nullity(T) + Rank(T) = n + 0 = n$.

Conversely, assume $dim(V) = dim(W)$. Chose bases $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ of V and $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ in W and define $T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots c_n\boldsymbol{v}_n) = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \dots + c_n\boldsymbol{w}_n$.

This is a **linear transformation**. Since $(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n)) = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_n)$, a **linearly independent sequence**, by **Theorem** (6.2.5) we can conclude that T is **one-to-one**. Since $R(T) = Span(T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n)) = Span(\boldsymbol{w}_1, \dots, \boldsymbol{w}_n) = W$ it also follows that T is **onto**. Thus, T is a **isomorphism**. $\square$

Example 6.2.5. (1) $\mathbb{R}_2[x]$ is **isomorphic** to $\mathbb{R}^3$. The **linear transformation** which takes $f(x) = a_0 + a_1x + a_2x^2$ to $\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$ is an **isomorphism**. So is the **linear tranformation** which takes $f \in \mathbb{R}_2[x]$ to $\begin{pmatrix} f(1) \\ f(2) \\ f(3) \end{pmatrix}$.

(2) $M_{m\times n}(\mathbb{R})$ is **isomorphic** to $\mathbb{R}^{mn}$. We illustrate with M_{23}. Define $T : M_{23} \to \mathbb{R}^6$ by

$$T(\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \\ a_{13} \\ a_{23} \end{pmatrix}.$$

The **linear transformation** T is an **isomorphism**.

What You Can Now Do

In all of the following we assume that V is a **finite dimensional vector space**.

1. Given a **linear transformation** $T : V \to W$ find a **basis** for the **range** of $T, R(T)$.

2. Given a **linear transformation** $T : V \to W$ determine the **rank** of $T, Rank(T)$.

3. Given a **linear transformation** $T : V \to W$ determine if T is **onto**.

4. Given a **linear transformation** $T : V \to W$ find a **basis** for the **kernel** of $T, Ker(T)$.

5. Given a **linear transformation** $T : V \to W$ determine the **nullity** of $T, Nullity(T)$.

6. Given a **linear transformation** $T : V \to W$ determine if T is **one-to-one**.

7. Given a **linear transformation** $T : V \to W$ determine if T is an **isomorphism**.

Method (How To Do It)

When we say that "We are given a **linear transformation** $T : V \to W$" this will either mean that we have a formula giving us the **image** of a typical vector or that we know the images of a **basis** for V.

Method 6.2.1. Given a **linear transformation** $T : V \to W$ find a **basis** for **range** of $T, R(T)$.

If we are given T by a formula then we can express the general **image vector** as a **linear combination** of some vectors and this exercise reduces to finding a **basis** for the span of those vectors. Of course, this may be going on in an arbitary vector space and not $\mathbb{R}^m$. If we are given $T(\boldsymbol{v}_1), \ldots, T(\boldsymbol{v}_n)$ for some **basis** $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of V then we have to find a basis for $Span(T(\boldsymbol{v}_1), \ldots, T(\boldsymbol{v}_n))$.

Example 6.2.6. Let $T : \mathbb{R}_2[x] \to M_{2\times 2}(\mathbb{R})$ be given by $T(a + bx + cx^2) = \begin{pmatrix} a+b+2c & a+3b-2c \\ a+2b & a+4b-4c \end{pmatrix}$. Find a **basis** for $R(T)$.

$$T(a + bx + cx^2) = a\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + b\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} + c\begin{pmatrix} 2 & -2 \\ 0 & -4 \end{pmatrix}.$$

To find a basis we "turn" the vectors into 4-vectors and find a **basis** for the **subspace** of $\mathbb{R}^4$ **spanned** by these columns. (Remark: We are "turning" the vectors into columns by using their **coordinate vectors** with respect to the **standard basis**). The relevant matrix is $\begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 0 \\ 1 & 3 & -2 \\ 1 & 4 & -4 \end{pmatrix}$.

This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

So the first and second columns are **pivot columns**. Thus,

$$R(T) = Span\left(\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}\right).$$

and $\left(\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}\right)$ is a **basis** for $R(T)$.

Example 6.2.7. Let $S : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}_4[x]$ be given by

$$S\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = 1 + x^2, S\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = -x + x^3,$$

$$S\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = 1 + x^3, S\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 1 + x^4.$$

We associate these to columns (by using their coordinate vectors with respect to the standard basis for $\mathbb{R}_4[x]$), make a matrix and obtain its **reduced echelon form** using **Gaussian elimination**. We can then use this to determine a subsequence which is **linearly independent** and therefore a **basis** for the **subspace** of $\mathbb{R}_4[x]$ which the sequence **spans**. The matrix is $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and therefore the four image vectors are linearly **linearly independent** and are a **basis** for the **range** of S.

Method 6.2.2. Given a **linear transformation** $T : V \to W$ determine the **rank** of $T, Rank(T)$.

Use **Method** (6.2.1) to obtain a **basis** for the **range** of T and count the number of basis elements to get the **dimension** of $R(T)$. This is the **rank** of T.

Example 6.2.8. For the **linear transformation** T of **Method** (6.2.6) we have $Rank(T) = 2$. For the **linear transformation** of **Example** (6.2.7), $Rank(S) = 4$.

Method 6.2.3. Given a **linear transformation** $T : V \to W$ determine if T is **onto**.

Use **Method** (6.2.2) to determine the **rank** of $T, Rank(T)$. T is **onto** when $Rank(T) = dim(W)$.

Example 6.2.9. Let $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}_2[x]$ be given by

$$T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a+2b+2c+d) + (2a+3b+c+d)x + (-3a-5b-3c+d)x^2.$$

Compute the **rank** of T and determine if T is **onto**.

$$T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a(1+2x-3x^2) + b(2+3x-5x^2) + c(2+x-3x^2) + d(1+x+x^2)$$

and therefore the **range** of T, $R(T)$ is the **span** of $(1+2x-3x^2), (2+3x-5x^2), (2+x-3x^2), (1+x+x^2)$. We make these into the columns of matrix (by taking their **coordinate vectors** with respect to the **standard basis** of $\mathbb{R}_2[x]$). We then use **Gaussian elimination** to obtain the **reduced echelon form**. We can then use this to find a **basis** for the **range**. The matrix is $\begin{pmatrix} 1 & 2 & 2 & 1 \\ 2 & 3 & 1 & 1 \\ -3 & -5 & -3 & 1 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -4 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

So the $1^{st}, 2^{nd}$ and 4^{th} columns are the **pivot columns** and therefore

$$R(T) = Span(1 + 2x - 3x^2, 2 + 3x - 5x^2, 1 + x + x^2).$$

Consequently, the **rank** of T, $Rank(T) = 3$ and T is **onto**.

Method 6.2.4. Given a **linear transformation** $T : V \rightarrow W$ find a **basis** for the **kernel** of T, $Ker(T)$.

This ultimately comes down to finding the **solution space** of a **homogeneous linear system**. We illustrate for each of the two ways in which T might be defined.

Example 6.2.10. Let $T : M_{2\times 3}(\mathbb{R}) \rightarrow M_{2\times 2}(\mathbb{R})$ be given by

$$T\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} a + b + 2c + f & -2a - 3b - c + e \\ a + 2b - c + d + 3e & -d - 4e - f \end{pmatrix}$$

Find a **basis** for the **kernel** of T, $Ker(T)$.

We need to determine when a matrix in $M_{2\times 3}(\mathbb{R})$ is mapped to the **zero vector** of $M_{2\times 2}(\mathbb{R})$. This gives rise to the **homogeneous linear system** with **coefficient matrix**

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

This matrix has **reduced echelon form**

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

The **homogeneous linear system** with **matrix** R has **leading variables** a, b, d and **free variables** c, e, f. We treat c, e, f as parameters and solve for a, b, d in terms of them. This is what we get

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{pmatrix} -5c-e-3f \\ 3c+e+2f \\ c \\ -4e-f \\ e \\ f \end{pmatrix} = c\begin{pmatrix} -5 \\ 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + e\begin{pmatrix} -1 \\ 1 \\ 0 \\ -4 \\ 1 \\ 0 \end{pmatrix} + f\begin{pmatrix} -3 \\ 2 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

From this we conclude that

$$Ker(T) = Span\left(\begin{pmatrix} -5 & 3 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 1 & 0 \\ -4 & 1 & 0 \end{pmatrix}, \begin{pmatrix} -3 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}\right)$$

Thus, $\left(\begin{pmatrix} -5 & 3 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 1 & 0 \\ 4 & 1 & 0 \end{pmatrix}, \begin{pmatrix} -3 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}\right)$ is the desired **basis**.

Example 6.2.11. Let $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}^3$ be given by

$$T\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, T\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 3 \end{pmatrix},$$

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

This implies that $T\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a+2b-2c+d \\ 2a+5b-4c+d \\ 2a+3b-4c+3d \end{pmatrix}$.

We have to find those vectors which have image $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ and so find the **solutions** to the **homogeneous linear system** with **coefficient matrix** $A = \begin{pmatrix} 1 & 2 & -2 & 1 \\ 2 & 5 & -4 & 1 \\ 2 & 3 & 4 & 3 \end{pmatrix}$.

This matrix A has **reduced echelon form** $R = \begin{pmatrix} 1 & -0 & 2 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. There are two **leading variables** a, b and two **free variables** c, d. Setting $c = s, d = t$ we get

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 2s-3t \\ t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Then $Ker(T) = Span\left(\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -3 & 0 \\ 1 & 1 \end{pmatrix}\right)$. The sequence, $\left(\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -3 & 0 \\ 1 & 1 \end{pmatrix}\right)$ is a **basis** for the **kernel** of T, $Ker(T)$.

Method 6.2.5. Given a **linear transformation** $T : V \to W$ determine the **nullity** of T, $Nullity(T)$.

Use **Method** (6.2.4) to find a **basis** for the **kernel** of T, $Ker(T)$. The length of this **basis** is the **dimension** of $Ker(T)$ which is the **nullity** of T, $Nullity(T)$.

Example 6.2.12. For the **linear transformation** of **Example** (6.2.10), $Nullity(T) = 3$. For the **linear transformation** of **Example** (6.2.11), $Nullity(T) = 2$.

Method 6.2.6. Given a **linear transformation** $T : V \to W$ determine if T is **one-to-one**.

T is **one-to-one** when $Ker(T) = \{\mathbf{0}_V\}$, equivalently, when $Nullity(T) = 0$. Use **Method** (6.2.4) to compute a **basis** for $Ker(T)$ and use this to determine if T is **one-to-one**.

Example 6.2.13. Let $T : M_{2\times 2}(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ be given by

$$T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+b+3c+d & 2a+3b+5c+2d \\ a+2b+3c & 2a+3b+4c+3d \end{pmatrix}.$$

Determine if this **linear transformation** is **one-to-one**.

Setting this equal to $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ we get the **homogeneous linear system** with **coeff-**

icient matrix $A = \begin{pmatrix} 1 & 1 & 3 & 1 \\ 1 & 2 & 3 & 0 \\ 2 & 3 & 5 & 2 \\ 2 & 3 & 4 & 3 \end{pmatrix}$.

The matrix A has reduced echelon form $R = \begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. There is one free variable so $Nullity(T) = 1$ and T is not one-to-one.

Example 6.2.14. Let $T : \mathbb{R}_2[x] \to \mathbb{R}^4$ be the linear transformation given by $T(1) = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 2 \end{pmatrix}$, $T(x) = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}$, $T(x^2) = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}$. Determine if T is one-to-one.

We have to determine if $(T(1), T(x), T(x^2))$ is linearly independent. Thus we make the matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$. This matrix has reduced echelon form $R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ and therefore T is one-to-one.

Method 6.2.7. Given a linear transformation $T : V \to W$ determine if T is an isomorphism.

If $dim(V) \neq dim(W)$ then we can immediately conclude that T is not an isomorphism. If $dim(V) = dim(W)$ then we have to check if the linear transformation is one-to-one or onto by Theorem 6.2.6).

Example 6.2.15. Let $T : \mathbb{R}^3 \to M_{2\times 2}(\mathbb{R})$ be defined by

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

Is T an isomorphism?

No since $dim(\mathbb{R}^3) = 3$ and $dim(M_{2\times 2})(\mathbb{R}) = 4$. Necessarily, this **linear transformation** cannot be **onto**.

Example 6.2.16. Let $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}_2[x]$ be given by

$$T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a + b - 2c) + (b + c - 2d)x + (-2a + c + d)x^2.$$

Is T an **isomorphism**?

No, $dim(M_{2\times 2}(\mathbb{R})) = 4$ while $dim(\mathbb{R}_2[x]) = 3$. Necessarily this **linear transformation** is not **one-to-one**.

Example 6.2.17. Let $T : \mathbb{R}^4 \to M_{2\times 2}(\mathbb{R})$ be given by

$$T(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}) = \begin{pmatrix} x_1 - x_2 & x_2 + x_3 - x_4 \\ x_2 - x_3 & 2x_1 - x_4 \end{pmatrix}.$$

Determine if T is an **isomorphism**.

Since $dim(\mathbb{R}^4) = 4 = dim(M_{22})$ it is possible that T is an **isomorphism**. We have to determine if $Ker(T) = \{\mathbf{0}_4\}$. Setting $T(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}) = \begin{pmatrix} x_1 - x_2 & x_2 + x_3 - x_4 \\ x_2 - x_3 & 2x_1 - x_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ we get the following **homogeneous linear system**:

$$\begin{array}{ccccccccc} x_1 & - & x_2 & & & & & = & 0 \\ & & x_2 & - & x_3 & & & = & 0 \\ & & x_2 & + & x_3 & - & x_4 & = & 0 \\ 2x_1 & & & & & + & x_4 & = & 0 \end{array}.$$

The **matrix** of this homogeneous linear system is $A = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ 2 & 0 & 0 & -1 \end{pmatrix}$.

The **reduced echelon form** of this matrix is $R = \begin{pmatrix} 1 & 0 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$. Therefore $Ker(T) \neq \{\mathbf{0}_4\}$ and T is not an **isomorphism**. In fact,

$$Ker(T) = Span\left(\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}\right) = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}\right).$$

Example 6.2.18. Let $T : \mathbb{R}_2[x] \to \mathbb{R}^3$ be given by

$$T(a + bx + cx^2) = \begin{pmatrix} a + 2b + 2c \\ a + 3b + c \\ 2a + 5b + 2c \end{pmatrix}.$$

Determine if T is an **isomorphism**.

Since $dim(\mathbb{R}_2[x]) = 3 = dim(\mathbb{R}^3)$ we have to determine if $Ker(T)$ is $\{\mathbf{0}_{\mathbb{R}_2[x]}\}$. Setting $\begin{pmatrix} a + 2b + 2c \\ a + 3b + c \\ 2a + 5b + 2c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ we have to find the **solutions** to the **homogeneous linear system** with **matrix** $A = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 5 & 2 \end{pmatrix}$.

The matrix A has **reduced echelon form** equal to the 3×3 **identity matrix** , I_3, and we conclude that T is an **isomorphism**.

Exercises

In 1- 7 find a **basis** for the **range** and the **rank** of the given **linear transformation** and determine if it is **onto**. See **Method** (6.2.1), **Method** (6.2.2), and **Method** (6.2.3).

1. $T : \mathbb{R}_3[x] \to \mathbb{R}^4$ given by

$$T(a + bx + cx^2 + dx^3) = \begin{pmatrix} a + 2b + 2d \\ a + 3b + c + d \\ a + b - c + d \\ a + 2b + 2d \end{pmatrix}$$

2. $T : \mathbb{R}_2[x] \to M_{2\times 2}(\mathbb{R})$ given by

$$T(1+x) = \begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix}, T(x+x^2) = \begin{pmatrix} 5 & 5 \\ 2 & 1 \end{pmatrix}, T(1+2x+2x^2) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

3. $T : \mathbb{R}^4 \to M_{2\times 2}(\mathbb{R})$ given by $T(\boldsymbol{e}_1) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, T(\boldsymbol{e}_2) = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, T(\boldsymbol{e}_3) = \begin{pmatrix} -1 & 2 \\ 1 & 4 \end{pmatrix}, T(\boldsymbol{e}_4) = \begin{pmatrix} 3 & 2 \\ 3 & 1 \end{pmatrix}$.

4. $T : \mathbb{R}_2[x] \to \mathbb{R}^4$ given by $T(1) = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix}, T(x) = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 3 \end{pmatrix}, T(x^2) = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 3 \end{pmatrix}$.

5. $T : \mathbb{R}_3[x] \to \mathbb{R}_2[x]$ given by $T(a + bx + cx^2 + dx^3) = (a + 2b + c) + (2a + 5b + c + d)x + (2a + 6b + d)x^2$.

6. $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}_3[x]$ given by
$T\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = 1 + x + x^2 + 2x^3, T\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 1 + 2x + 3x^2 + x^3,$

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

7. $T : \mathbb{R}_4[x] \to \mathbb{R}^4$ given by

$$T(1) = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, T(x) = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 3 \end{pmatrix}, T(x^2) = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, T(x^3) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 2 \end{pmatrix}, T(x^4) = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 2 \end{pmatrix}.$$

In exercises 8 - 14 find a **basis** for the **kernel** and determine the nullity of the given **linear transformation**. Then determine if the **linear transformation** is **one-to-one**. See **Method** (6.2.4), **Method** (6.2.5), and **Method** (6.2.6).

8. The **linear transformation** of exercise 1.

9. The **linear transformation** of exercise 2.

10. The **linear transformation** of exercise 3.

11. The **linear transformation** of exercise 4.

12. The **linear transformation** of exercise 5.

13. The **linear transformation** of exercise 6.

14. The **linear transformation** of exercise 7.

In exercises 15 - 21 determine if the **linear transformation** is an **isomorphism**. See **Method** (6.2.7).

15. The **linear transformation** of exercise 1.

16. The **linear transformation** of exercise 2.

17. The **linear transformation** of exercise 3.

18. The **linear transformation** of exercise 4.

19. The **linear transformation** of exercise 5.

20. The **linear transformation** of exercise 6.

21. The **linear transformation** of exercise 7.

In exercises 22 - 25 answer true or false and give an explanation.

22. If $T : V \to W$ is a **linear transformation** of **finite dimensional vector spaces** then the **rank** of T is positive, $Rank(T) > 0$.

23. If $T : V \to W$ is a **linear transformation** of **finite dimensional vector spaces** and $Rank(T) = 3$ then the $Ker(T)$, has a **basis** with three elements.

24. If $T : \mathbb{R}^7 \to \mathbb{R}^3$ is a **linear transformation** and **onto** then the $Nullity(T)$ = 3.

25. If $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}_3[x]$ then T cannot be an **isomorphism**.

Challenge Exercises (Problems)

In 1 and 2 let $S : U \to V$ and $T : V \to W$ be transformations (not necessarily linear) and let $T \circ S : U \to W$ be the composition given by $(T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u}))$.

1. Assume that S and T are both **onto**. Prove that $T \circ S$ is onto

2. Assume that S and T are both **one-to-one**. Prove that $T \circ S$ is one-to-one.

3. Let V and W be **finite dimensional vector spaces** and $T : V \to W$ a **linear transformation**. Prove that if T is **onto** then $dim(V) \geq dim(W)$.

4. Let V and W be **finite dimensional vector spaces** and $T : V \to W$ a **linear transformation** . Prove that if T is **one-to-one** then $dim(V) \leq dim(W)$.

5. Let $T : M_{2\times 3}(\mathbb{R}) \to \mathbb{R}_4[x]$ be a **linear transformation** and assume that the following sequence of vectors is a **basis** for $R(T) : (1 + x^2 + x^4, x + x^3, 1 + x + 2x^2)$.

What is the **Rank** and **Nullity** of T?

Quiz Solutions

1. $\begin{pmatrix} 1 & 2 & 1 & 1 & 0 \\ 1 & 3 & 0 & 2 & 1 \\ 2 & 5 & 1 & 4 & 1 \\ 3 & 5 & 4 & 2 & -1 \end{pmatrix}$

The **reduced echelon form** of the matrix is $\begin{pmatrix} 1 & 0 & 3 & 0 & -2 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

The **pivot columns** are 1,2,4 and therefore the sequence $\left(\begin{pmatrix} 1 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 4 \\ 2 \end{pmatrix}\right)$ is a **basis** for the **column space**.

There are two **non-pivot columns** (3 and 5) and so two **free variables** (x_3, x_5). We set these equal to parameters, s and t and solve for the other variables:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -3s+2t \\ s-t \\ s \\ 0 \\ t \end{pmatrix} = s\begin{pmatrix} -3 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Thus a **basis** for the **null space** is $\left(\begin{pmatrix} -3 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right)$.

The **rank** is 3 and the **nullity** is 2. A **basis** for the **row space** is

$$\left(\begin{pmatrix} 1 \\ 0 \\ 3 \\ 0 \\ -2 \end{pmatrix}^{Tr}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}^{Tr}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}^{Tr}\right).$$

The sequence of columns do not **span** $\mathbb{R}^4$ and is **linearly dependent**.

Not right, **Method** (5.5.1) and **Method** (5.5.2).

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

The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Columns 1,2 4 are a **basis** for the **column space** which is $\mathbb{R}^3$. The vector $\begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ is a **basis** for the **null space**. The sequence of columns is **linearly dependent**. The **rank** is 3 and the **nullity** is 1.

The sequence $\left(\begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix}^{Tr}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}^{Tr}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}^{Tr} \right)$ is a **basis** for the **row space**.

Not right, **Method** (5.5.1) and **Method** (5.5.2).

6.3. Matrix of a Linear Transformation

In this section we show that a **linear transformation** between two **finite dimensional vector spaces** can be represented by **matrix multiplication** and then show how this can be used to compute a **basis** for the **range** and for the **kernel** of the **linear transformation**. In turn, we can use this to compute the **rank** and **nullity** of a linear transformation.

Am I Ready for This Material

Readiness Quiz

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

The following are concepts that are used extensively in this section:

vector space

basis of a vector space

coordinate vector with respect to a basis

change of basis matrix

linear transformation

range of a transformation

kernel of a linear transformaion

standard basis of $\mathbb{R}^n$

matrix transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$

standard matrix of a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$

standard basis of $\mathbb{R}_n[x]$

standard basis of $M_{m\times n}(\mathbb{R})$

Quiz

1. Compute the **product** $\begin{pmatrix} 2 & 4 & 7 & -5 \\ 3 & -6 & 9 & 2 \\ -4 & 1 & 8 & -3 \end{pmatrix} \begin{pmatrix} 3 \\ -5 \\ 7 \\ 2 \end{pmatrix}$.

2. Let $A = \begin{pmatrix} 1 & 3 & 2 & 2 & -1 \\ 2 & 5 & 5 & 3 & -4 \\ 2 & 7 & 4 & 1 & -3 \end{pmatrix}$. Find the **null space** of A.

3. Let $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 4 \\ 4 & 3 & 3 & 5 \\ 3 & 2 & 1 & 1 \\ 7 & 5 & 4 & 6 \end{pmatrix}$. Determine if $\begin{pmatrix} 1 \\ 8 \\ 7 \\ -3 \\ 4 \end{pmatrix}$ is in the **column space** of A.

Quiz Solutions

New Concepts

Two very important concepts are introduced here:

matrix of a linear transformation

similarity of matrices

Theory (Why It Works)

Let $T : V \to W$ be a **linear transformation** from an **n-dimensional vector space** V to an $m-$dimensinal vector space W. Let $\mathcal{B}_V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a **basis** for V and $\mathcal{B}_W = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be a basis for W.

Then each $T(\boldsymbol{v}_j)$ can be written in a unique way as a **linear combination** of the **basis** $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m)$:

$T(\boldsymbol{v}_j) = a_{1j}\boldsymbol{w}_1 + a_{2j}\boldsymbol{w}_2 + \cdots + a_{mj}\boldsymbol{w}_m$ which is the same thing as

$$[T(\boldsymbol{v}_j)]_{\mathcal{B}_W} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Let A be the $m \times n$ matrix with entries $a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n$ so that, $A = ([T(\boldsymbol{v}_1)]_{\mathcal{B}_W}\ [T(\boldsymbol{v}_2)]_{\mathcal{B}_W}\ \ldots\ [T(\boldsymbol{v}_n)]_{\mathcal{B}_W})$. Now suppose $\boldsymbol{v} \in V$ and the **coordinate vector** of $\boldsymbol{v}$ with respect to $\mathcal{B}_V$ is $[\boldsymbol{v}]_{\mathcal{B}_V} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$. This means that

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$$

is the unique expression of $\boldsymbol{v}$ as a **linear combination** of the **basis** $\mathcal{B}_V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$.

As a consequence of the fact that T is a **linear transformation** we have

$$T(\boldsymbol{v}) = T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \ldots c_nT(\boldsymbol{v}_n) =$$

$$c_1(a_{11}\boldsymbol{w}_1 + a_{21}\boldsymbol{w}_2 + \cdots + a_{m1}\boldsymbol{w}_m) + c_2(a_{12}\boldsymbol{w}_1 + a_{22}\boldsymbol{w}_2 + \cdots + a_{m2}\boldsymbol{w}_m) + \cdots +$$

$$c_n(a_{1n}\boldsymbol{w}_1 + a_{2n}\boldsymbol{w}_2 + \cdots + a_{mn}\boldsymbol{w}_m) =$$

$$(c_1a_{11} + c_2a_{12} + \cdots + c_na_{1n})\boldsymbol{w}_1 + (c_1a_{21} + c_2a_{22} + \cdots + c_na_{2n})\boldsymbol{w}_2 + \cdots +$$

$$(c_1a_{m1} + c_2a_{m2} + \cdots + c_na_{mn})\boldsymbol{w}_m.$$

Therefore

$$[T(\boldsymbol{v})]_{\mathcal{B}_W} = \begin{pmatrix} c_1a_{11} + c_2a_{12} + \cdots + c_na_{1n} \\ c_1a_{21} + c_2a_{22} + \cdots + c_na_{2n} \\ \vdots \\ c_1a_{m1} + c_2a_{m2} + \cdots + c_na_{mn} \end{pmatrix} =$$

$$c_1\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + c_2\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + c_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

But this means that $[T(\boldsymbol{v})]_{\mathcal{B}_W} = A[\boldsymbol{v}]_{\mathcal{B}_V}$. We have therefore established most of the following result:

Theorem 6.3.1. *Let $T : V \to W$ be a* **linear transformation** *between two* **finite dimensional vector spaces** *V and W. Let $\mathcal{B}_V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a* **basis** *for V and $\mathcal{B}_W = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be a basis for W. Let A be the matrix whose columns are $[T(\boldsymbol{v}_j)]_{\mathcal{B}_W}, j = 1, 2, \ldots, n$. Then A satisfies: $[T(\boldsymbol{v}_j)]_{\mathcal{B}_W} = A[v_j]_{\mathcal{B}_V}$. Moreover, A is the only matrix with this property.*

Proof. We have proved everything except the uniqueness of the matrix A. So suppose A' is an $m \times n$ matrix and also

$$[T(\boldsymbol{v})]_{\mathcal{B}_W} = A'[\boldsymbol{v}]_{\mathcal{B}_V}. \tag{6.6}$$

We need to show that $A' = A$. Since (6.6) holds for all vectors $\boldsymbol{v} \in V$, in particular, it is satisfied for $\boldsymbol{v}_j$. Note that $[\boldsymbol{v}_j]_{\mathcal{B}_V} = \boldsymbol{e}_j^n$ the j^{th} **standard basis vector** in $\mathbb{R}^n$. Thus, we have

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} = [T(\boldsymbol{v}_j)]_{\mathcal{B}_W} = A'[\boldsymbol{v}_j]_{\mathcal{B}_V} = A'\boldsymbol{e}_j^n. \tag{6.7}$$

If follows from (6.7) that the j^{th} column of A' is $[T(\boldsymbol{v}_j)]_{\mathcal{B}_W}$ for each $j, 1 \leq j \leq n$. However, this is also the j^{th} column of A and therefore $A' = A$ as required. $\square$

Theorem (6.3.1) inspires the following definition:

Definition 6.11. Let $T : V \to W$ be a **linear transformation** between two **finite dimensional vector spaces** V and W. Let $\mathcal{B}_V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a basis for V and $\mathcal{B}_W = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be a basis for W. Let $\mathcal{M}_T(\mathcal{B}_V, \mathcal{B}_W)$ denote the matrix whose columns are $[T(\boldsymbol{v}_j)]_{\mathcal{B}_W}, j = 1, 2, \ldots, n$. The matrix $\mathcal{M}_T(\mathcal{B}_V, \mathcal{B}_W)$ is called the ***matrix of*** T with respect to the **bases** $\mathcal{B}_V$ and $\mathcal{B}_W$.

Example 6.3.1. Assume that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation**, $\mathcal{S}_n = (\boldsymbol{e}_1^n, \ldots, \boldsymbol{e}_n^n)$ is the **standard basis** of $\mathbb{R}^n$ and $\mathcal{S}_m = (\boldsymbol{e}_1^m, \ldots, \boldsymbol{e}_m^m)$ is the standard basis of $\mathbb{R}^m$ and set $A = \mathcal{M}_T(\mathcal{S}_n, \mathcal{S}_m)$. Recall, we previously referred to this matrix as the **standard matrix of T**. We do so because it is the matrix of T with respect to the **standard bases** of the respective spaces.

Example 6.3.2. Let V be a **finite dimensional vector space** with bases $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}_1', \ldots, \boldsymbol{v}_n')$. Consider the identity transformation $I_V : V \to V$ which takes a vector $\boldsymbol{v}$ to itself: $I_V(\boldsymbol{v}) = \boldsymbol{v}$. What is $\mathcal{M}_{I_V}(\mathcal{B}, \mathcal{B}')$? The j^{th} column of this matrix is the **coordinate vector** of $I_V(\boldsymbol{v}_j) = \boldsymbol{v}_j$ with respect to the **basis** $\mathcal{B}'$. We have encountered this matrix before: it is the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$, $\mathcal{M}_{I_V}(\mathcal{B}, \mathcal{B}') = P_{\mathcal{B} \to \mathcal{B}'}$.

Example 6.3.3. Define $T : \mathbb{R}_3[x] \to \mathbb{R}^4$ by $T(\boldsymbol{f}) = \begin{pmatrix} \boldsymbol{f}(1) \\ \boldsymbol{f}(2) \\ \boldsymbol{f}(3) \\ \boldsymbol{f}(4) \end{pmatrix}$. Find the **matrix** of T with respect to the **standard basis** of $\mathbb{R}_3[x]$, $\mathcal{S}_{\mathbb{R}_3[x]} = (1, x, x^2, x^3)$, and the **standard basis** $\mathcal{S}_4$ of $\mathbb{R}^4$.

We compute $T(1), T(x), T(x^2), T(x^3)$:

$$T(1) = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, T(x) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix},$$

$$T(x^2) = \begin{pmatrix} 1 \\ 4 \\ 9 \\ 16 \end{pmatrix}, T(x^3) = \begin{pmatrix} 1 \\ 8 \\ 27 \\ 64 \end{pmatrix}.$$

Therefore the matrix of the transformation is $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{pmatrix}$.

When the **domain and codomain** of a **linear transformation** are the same **vector space**, then it is customary (but not necessary) to use the same **basis** for both. We illustrate with an example but first, for future reference, we give a name to **linear transformations** in which the domain and the codomain are the same **vector space**.

Definition 6.12. Assume V is a **vector space**. A **linear transformation** $T : V \to V$ is referred to as a ***linear operator***.

Example 6.3.4. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the **matrix transformation** with **standard matrix** $A = \begin{pmatrix} 2 & 0 & -1 \\ -1 & 1 & 1 \\ -2 & 0 & 3 \end{pmatrix}$. Consider the sequence $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} \right)$.

We verify that $\mathcal{B}$ is a **basis** of $\mathbb{R}^3$. By **Theorem** (2.4.4) we need only show that $\mathcal{B}$ is **linearly independent**. This is demonstrated by forming the matrix P with the vectors of $\mathcal{B}$ as columns, applying **Gaussian elimination** to show that the **reduced echelon form** of P is the 3×3 **identity matrix** , I_3. Since we will want the **inverse** of P and it is not much more effort to compute P^{-1} we will do this in the course of finding the **reduced echelon form** of $P = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}$.

We augment the **identity matrix** to P and apply **Gaussian elimination**:

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

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

So, indeed, the columns are **linearly independent** and $\mathcal{B}$ is a **basis** of $\mathbb{R}^3$. We now compute the **images** of the vectors of $\mathcal{B}$:

$$T(\begin{pmatrix}1\\0\\1\end{pmatrix}) = \begin{pmatrix}1\\0\\1\end{pmatrix}, T(\begin{pmatrix}0\\1\\0\end{pmatrix}) = \begin{pmatrix}0\\1\\0\end{pmatrix}, T(\begin{pmatrix}-1\\1\\2\end{pmatrix}) = \begin{pmatrix}-4\\4\\8\end{pmatrix} = 4\begin{pmatrix}-1\\1\\2\end{pmatrix}.$$

Thus, the **matrix** $B = \mathcal{M}_T(\mathcal{B}, \mathcal{B})$ of T with respect to the basis $\mathcal{B}$ is $\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 4\end{pmatrix}$.

Remark 6.2. When we study eigenvalues and eigenvectors we will better understand this example.

There are two things to observe about **Example** (6.3.4):

1. The matrix P is the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{S}$, the **standard basis** of $\mathbb{R}^3$, and P^{-1} is the change of basis matrix from $\mathcal{S}$ to $\mathcal{B}$.

2. $B = P^{-1}AP$.

We could multiply the matrices to verify this but we can see this more elegantly and in this way preview the general argument.

We have shown that

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

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

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

Consequently.

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

By multiplying both sides by $P^{-1} = \begin{pmatrix} \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & 1 & -\frac{1}{3} \\ -\frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$ we obtain the relation $B = P^{-1}AP$.

This phenonmenon happens in general.

Theorem 6.3.2. *Let V be a **finite dimensional vector space** with **bases** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$.*

*Let $T : V \to V$ be a **linear transformation** and assume the **matrix** of T with respect to $\mathcal{B}$ is A and the matrix of T with respect to $\mathcal{B}'$ is A'. Let $P = P_{\mathcal{B}' \to \mathcal{B}}$ be the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$. Then $A' = P^{-1}AP$.*

Proof. Because P is the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$ the inverse, P^{-1}, is the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$. Suppose $\boldsymbol{w}$ is a vector in V. Then $[\boldsymbol{w}]_{\mathcal{B}'} = P^{-1}[\boldsymbol{w}]_{\mathcal{B}}$. In particular, for a vector $\boldsymbol{v}$ in V, $[T(\boldsymbol{v})]_{\mathcal{B}'} = P^{-1}[T(\boldsymbol{v})]_{\mathcal{B}}$. Since A is the **matrix** of T with respect to $\mathcal{B}$: $[T(\boldsymbol{v})]_{\mathcal{B}} = A[\boldsymbol{v}]_{\mathcal{B}}$. So, making this substitution we have $[T(\boldsymbol{v})]_{\mathcal{B}'} = P^{-1}(A[\boldsymbol{v}]_{\mathcal{B}})$.

Since P is the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$ we have $[\boldsymbol{v}]_{\mathcal{B}} = P[\boldsymbol{v}]_{\mathcal{B}'}$. Making this substitution we get $[T(\boldsymbol{v})]_{\mathcal{B}'} = P^{-1}(A[\boldsymbol{v}]_{\mathcal{B}}) = P^{-1}(PA[\boldsymbol{v}]_{\mathcal{B}'}) = (P^{-1}AP)[\boldsymbol{v}]_{\mathcal{B}'}$. By the definition of the **matrix** of T with respect to the basis $\mathcal{B}'$ it follows that $A' = P^{-1}AP$. □

The way in which the matrices of a **linear transformation** $T : V \to V$ which arise from different **bases** $\mathcal{B}$ and $\mathcal{B}'$ are related is fundamental and is the subject of the next definition:

Definition 6.13. Let A and B be $n \times n$ matrices. We say that A and B are ***similar*** if there exists an **invertible matrix** P such that $B = P^{-1}AP$.

Remark 6.3. 1. If A is an $n \times n$ matrix then A is **similar** to A. This follows by taking $P = I_n$.

2. If the $n \times n$ matrix B is **similar** to the $n \times n$ matrix A then A is similar to B.

Suppose that $B = P^{-1}AP$ for some **invertible matrix** P. Multiplying on the left hand side by P and the right hand side by P^{-1} we get $A = PBP^{-1}$. Since $(P^{-1})^{-1} = P$, if we set $Q = P^{-1}$ then $A = Q^{-1}BQ$.

3. If the $n \times n$ matrix B is **similar** to the $n \times n$ matrix A and the $n \times n$ matrix C is similar to B then C is similar to A.

B **similar** to A means that there is an **invertible matrix** P such that $B = P^{-1}AP$. C similar to B means that there is an invertible matrix Q such that $C = Q^{-1}BQ$. But then $C = Q^{-1}BQ = Q^{-1}(P^{-1}AP)Q = (PQ)^{-1}A(PQ)$. Since $(PQ)^{-1} = Q^{-1}P^{-1}$.

Another way of formulating **Theorem** (6.3.2) is as follows:

Theorem 6.3.3. *Let V be a* **n-dimensional vector space** *with* **bases** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ *and* $\mathcal{B}' = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$. *Let $T : V \to V$ be a* **linear transformation**, *A be the* **matrix** *of T with respect $\mathcal{B}$, and A' be the matrix of T with respect to $\mathcal{B}'$. Then A and A' are similar by the change of matrix $P = P_{\mathcal{B}' \to \mathcal{B}}$.*

Remark 6.4. The converse to **Theorem** (6.3.3) holds, that is, if two matrices A and A' are **similar** then they can be realized as matrices of the same **linear transformation** with respect to different **bases**.

Determinants of Linear Transformations

One consequence of **Theorem** (6.3.2) and **Theorem** (6.3.3) is that they provide a means of defining a determinant for a **linear operator** on a **vector space** V (a **linear transformation** T from a **finite dimensional vector space** V to itself).

Definition 6.14. Let $T : V \to V$ be a **linear operator** on the **finite dimensional vector space** V. Choose any **basis** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ for V and let $A = \mathcal{M}_T(\mathcal{B}, \mathcal{B})$ be the **matrix** of T with respect to $\mathcal{B}$. Define the ***determinant of*** T, denoted by $det(T)$, to be equal to $det(A)$.

It might appear that the definition of the **determinant** of T depends on the choice of a **basis** . This is not so. Let $\mathcal{B}' = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ be any other **basis** of V. Let P be the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$, $P = P_{\mathcal{B}' \to \mathcal{B}}$ and $B = \mathcal{M}_T(\mathcal{B}', \mathcal{B}')$ be the **matrix** of T with respect to $\mathcal{B}'$. Then by **Theorem** (6.3.2), $B = P^{-1}AP$ and $det(B) = det(P^{-1})det(A)det(P)$. Since $det(P^{-1})det(P) = 1$ we conclude that $det(B) = det(A)$ and so $det(T)$ is well-defined.

We now go on to prove some powerful results

Theorem 6.3.4. *Let $T : V \to W$ be a* **linear transformation** *between two* **finite dimensional vector spaces**. *Let $\mathcal{B}_V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a* **basis** *for V and $\mathcal{B}_W = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be a basis for W. Let $A = \mathcal{M}_T(\mathcal{B}_V, \mathcal{B}_W)$ be the* **matrix** *of T with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$. Then the following hold:*

1) A vector $\boldsymbol{v}$ is in the **kernel** *of T if and only if $[\boldsymbol{v}]_{\mathcal{B}_V}$ is in the* **null space** *of A.*

2) A vector $\boldsymbol{w}$ is in the **range** *of T if and only if $[\boldsymbol{w}]_{\mathcal{B}_W}$ is in the* **column space** *of A.*

Proof. 1) A vector $\boldsymbol{v}$ is in $Ker(T)$ if and only if $T(\boldsymbol{v}) = \boldsymbol{0}_W$ if and only if $[T(\boldsymbol{v})]_{\mathcal{B}_W} = \boldsymbol{0}_m \in \mathbb{R}^m$. However, $[T(\boldsymbol{v})]_{\mathcal{B}_W} = A[\boldsymbol{v}]_{\mathcal{B}_V}$ and therefore $\boldsymbol{v} \in Ker(T)$ if and only if $A[\boldsymbol{v}]_{\mathcal{B}_V} = \boldsymbol{0}_m \in \mathbb{R}^m$ if and only if $[\boldsymbol{v}]_{\mathcal{B}_V} \in null(A)$.

2) A vector $\boldsymbol{w}$ is in the **range** of T if and only if there is a vector $\boldsymbol{v}$ in V such that $T(\boldsymbol{v}) = \boldsymbol{w}$. Suppose that $\boldsymbol{w} \in R(T)$ and $\boldsymbol{w} = T(\boldsymbol{v})$. Then $[\boldsymbol{w}]_{\mathcal{B}_W} = [T(\boldsymbol{v})]_{\mathcal{B}_W} = A[\boldsymbol{v}]_{\mathcal{B}_V}$. Thus, $[\boldsymbol{w}]_{\mathcal{B}_W}$ is in the **column space** of A.

On the other hand, suppose that $[\boldsymbol{w}]_{\mathcal{B}_W}$ is in the **column space** of A. Then there is a vector $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ in $\mathbb{R}^n$ such that $[\boldsymbol{w}]_{\mathcal{B}_W} = A\boldsymbol{u}$. Let $\boldsymbol{v}$ be the vector in V such that $[\boldsymbol{v}]_{\mathcal{B}_V} = \boldsymbol{u}$, that is, $\boldsymbol{v} = u_1\boldsymbol{v}_1 + u_2\boldsymbol{v}_2 + \cdots + u_n\boldsymbol{v}_n$. Then $[T(\boldsymbol{v})]_{\mathcal{B}_W} = A[\boldsymbol{v}]_{\mathcal{B}_V} = A\boldsymbol{u} = [\boldsymbol{w}]_{\mathcal{B}_W}$ and, consequently, $\boldsymbol{w} = T(\boldsymbol{v})$. $\square$

An immediate consequence of **Theorem** (6.3.4) is that we can tell whether a **linear transformation** T is **one-to-one**, **onto**, or an **isomorphism** from examining the matrix $A = \mathcal{M}_T(\mathcal{B}_V, \mathcal{B}_W)$.

Theorem 6.3.5. *Let $T : V \to W$ be a* **linear transformation** *between two* **finite dimensional vector spaces**. *Let $\mathcal{B}_V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a* **basis** *for V and $\mathcal{B}_W = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be a basis for W. Let A be the matrix of T with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$, $A = \mathcal{M}_T(\mathcal{B}_V, \mathcal{B}_W)$. Then the following hold:*

1. T is **one-to-one** *if and only if $null(A) = \{\boldsymbol{0}_n\}$ if and only if A has n* **pivot columns** *if and only if the* **rank** *of A is n.*

2. T is **onto** *if and only if $col(A) = \mathbb{R}^m$ if and only if A has m* **pivot columns** *if and only if the* **rank** *of A is m.*

3. T is an **isomorphism** *if and only if A is an* **invertible matrix**.

What You Can Now Do

In all of the following V and W are assumed to be **finite dimensional vector spaces**.

1. Given a **linear transformation** $T : V \to W$ write down the **matrix** of the transformation with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W.

2. Given the **matrix** of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and and $\mathcal{B}_W$ for W and a vector $\boldsymbol{v} \in V$ compute the **image** $T(\boldsymbol{v})$.

3. Given the **matrix** A of a **linear transformation** $T : V \to V$ with respect to a **basis** $\mathcal{B}$ for V and given a second basis $\mathcal{B}'$ for V compute the matrix of T with respect to $\mathcal{B}'$.

4. Given the **matrix** of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W and the **coordinate vector** of $\boldsymbol{v} \in V$ determine if $\boldsymbol{v}$ is in the **kernel** of T.

5. Given the **matrix** of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W and the **coordinate vector** of $\boldsymbol{w} \in W$ determine if $\boldsymbol{w}$ is in the **range** of T.

6. Given the **matrix** of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W determine if T is **onto**, **one-to-one**, or an **isomorphism**.

Method (How To Do It)

Method 6.3.1. Given a **linear transformation** $T : V \to W$ write down the **matrix** of the transformation with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W.

"Given a **linear transformation**" either means that we are given a formula for computing the **image** of the typical element of the **domain** or else we are given the images of T on a **basis** of V. In the latter case this quickly leads to the matrix (provided the **basis** is the one to be used). In the former case we use the formula to compute the **images** of the **basis** vectors.

If $\mathcal{B}_V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ and $\mathcal{B}_W = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_m)$ then compute $T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \dots, T(\boldsymbol{v}_n)$.

Following this, compute the **coordinate vectors** of $T(\boldsymbol{v}_1), \dots, T(\boldsymbol{v}_n)$ respect to the basis $\mathcal{B}_W$, that is, compute $[T(\boldsymbol{v}_1)]_{\mathcal{B}_W}, [T(\boldsymbol{v}_2)]_{\mathcal{B}_W}, \dots, [T(\boldsymbol{v}_n)]_{\mathcal{B}_W}$.

The **matrix** of T with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$ is the matrix with the **coordinate vectors** $([T(\boldsymbol{v}_1)]_{\mathcal{B}_W}, [T(\boldsymbol{v}_2)]_{\mathcal{B}_W}, \dots, [T(\boldsymbol{v}_n)]_{\mathcal{B}_W})$ as its columns.

Example 6.3.5. Let $T : \mathbb{R}_2[x] \to M_{2\times 2}(\mathbb{R})$ be given by

$$T(a + bx + cx^2) = \begin{pmatrix} a - 2b + c & a - 3b + c \\ a - 3b + 2c & a + b + c \end{pmatrix}$$

a) Compute the **matrix** of T with respect to the **standard base** $\mathcal{S}_{\mathbb{R}_2[x]} = (1, x, x^2)$ of $\mathbb{R}_2[x]$ and the **standard basis** $\mathcal{S}_{M_{2\times 2}(\mathbb{R})} = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$ of $M_{2\times 2}(\mathbb{R})$.

b) Compute the **matrix** of T with respect to the **bases** $\mathcal{B}_{\mathbb{R}_2[x]} = (1 + x, x + x^2, 1 + 2x + 2x^2)$ and $\mathcal{B}_{M_{2\times 2}(\mathbb{R})} = \left(\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 4 \\ 3 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix} \right)$.

a) We compute the **images** of the **standard basis vectors**:

$$T(1) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, T(x) = \begin{pmatrix} -2 & -3 \\ -3 & 1 \end{pmatrix}, T(x^2) = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}.$$

The **coordinate vectors** of these images are

$$[T(1)]_{\mathcal{S}_{M_{2\times 2}(\mathbb{R})}} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$[T(x)]_{\mathcal{S}_{M_{2\times 2}(\mathbb{R})}} = \begin{pmatrix} -2 \\ -3 \\ -3 \\ 1 \end{pmatrix}$$

$$[T(x^2)]_{\mathcal{S}_{M_{2\times 2}(\mathbb{R})}} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}$$

Consequently, the **matrix** of the T with respect to $\mathcal{S}_{\mathbb{R}_2[x]}$ and $\mathcal{S}_{M_{2\times 2}(\mathbb{R})}$ is

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

b) The **images** of the vector in $\mathcal{B}_{\mathbb{R}_2[x]}$ are

$$T(1+x) = \begin{pmatrix} -1 & -2 \\ -2 & 2 \end{pmatrix}, T(x+x^2) = \begin{pmatrix} -1 & -2 \\ -1 & 2 \end{pmatrix}, T(1+2x+2x^2) = \begin{pmatrix} -1 & -3 \\ -1 & 5 \end{pmatrix}.$$

In order to get the matrix we have to find the **coordinate vectors** of these with respect to the **basis** $\mathcal{B}_{M_{22}}$. We use **Method** (5.4.1): We make an augmented matrix. On the left hand side we put the **coordinate vectors** of the vectors in $\mathcal{B}_{M_{2\times 2}(\mathbb{R})}$ with respect to the **standard basis** of $M_{2\times 2}(\mathbb{R})$; on the right hand side we put the **coordinate vectors** of the above images with respect to the standard base of $M_{2\times 2}(\mathbb{R})$. We use **Gaussian elimination** to get the **reduced echelon form**. On the left hand side we will get the 4×4 **identity matrix**, I_4 and on the right hand side the matrix of T with respect to the **bases** $\mathcal{B}_{\mathbb{R}_2[x]}$ and $\mathcal{B}_{M_{2\times 2}(\mathbb{R})}$.

The required matrix is $\begin{pmatrix} 1 & 2 & 2 & 0 & | & -1 & -1 & -1 \\ 1 & 3 & 1 & 2 & | & -2 & -1 & -1 \\ 1 & 4 & 1 & 3 & | & -2 & -2 & -3 \\ 0 & 1 & 1 & 1 & | & 2 & 2 & 5 \end{pmatrix}$.

The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 0 & 0 & | & -3 & 3 & 7 \\ 0 & 1 & 0 & 0 & | & -1 & -5 & -11 \\ 0 & 0 & 1 & 0 & | & 2 & 3 & 7 \\ 0 & 0 & 0 & 1 & | & 1 & 4 & 9 \end{pmatrix}$.

Therefore the **matrix** of T with respect to $\mathcal{B}_{\mathbb{R}_2[x]}$ and $\mathcal{B}_{M_{2\times 2}(\mathbb{R})}$ is $\begin{pmatrix} -3 & 3 & 7 \\ -1 & -5 & -11 \\ 2 & 3 & 7 \\ 1 & 4 & 9 \end{pmatrix}$.

Example 6.3.6. Let $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}^3$ be the **linear transformation** such that

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

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

Write the matrix of this transformation with respect to the **standard basis** of $M_{2\times 2}(\mathbb{R})$ and the **standard basis** of $\mathbb{R}^3$.

Since we are given the **images** of the **standard basis vectors** of $M_{2\times 2}(\mathbb{R})$ all we need do is write down the **coordinate vectors** of these images and then make the matrix with these as its columns. Thus, the **matrix** of T with respect to the standard bases of $M_{2\times 2}(\mathbb{R})$ and $\mathbb{R}^3$ is $\begin{pmatrix} 1 & -3 & -1 & 0 \\ -2 & 5 & 1 & -1 \\ 2 & -4 & 0 & 2 \end{pmatrix}$.

Method 6.3.2. Given the **matrix** of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and and $\mathcal{B}_W$ for W and a vector $\boldsymbol{v} \in V$ compute the **image** $T(\boldsymbol{v})$.

If we know the **coordinate vector** of $\boldsymbol{v}$ we can use the **matrix** of T to find the **coordinate vector** of its **image**: By **Theorem** (6.3.1 $[T(\boldsymbol{v})]_{\mathcal{B}_W} = A[\boldsymbol{v}]_{\mathcal{B}_V}$.

If $\mathcal{B}_W = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ and $[T(\boldsymbol{v})]_{\mathcal{B}_W} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$ then

$$T(\boldsymbol{v}) = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \cdots + c_m\boldsymbol{w}_m.$$

Example 6.3.7. Let $T : \mathbb{R}_3[x] \to M_{2\times 2}(\mathbb{R})$ be a **linear transformation** and assume that the **matrix** of T with respect to the **standard basis** of $\mathbb{R}_3[x]$ and the **standard basis** of $M_{2\times 2}(\mathbb{R})$ is

$A = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 2 & 3 & 1 & 3 \\ 1 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 \end{pmatrix}$. Find the image of $\boldsymbol{f} = 1 - 2x + 2x^2 + 4x^3$.

$$[T(\boldsymbol{f})]_{\mathcal{S}_{M_{2\times 2}(\mathbb{R})}} = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 2 & 3 & 1 & 3 \\ 1 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 10 \\ 7 \\ 6 \end{pmatrix}$$

Therefore $T(\boldsymbol{f}) = \begin{pmatrix} 1 & 7 \\ 10 & 6 \end{pmatrix}$.

Example 6.3.8. Let $T : \mathbb{R}^4 \to \mathbb{R}_3[x]$ be the linear transformation with matrix $\begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 3 & 1 & 0 \\ 2 & 3 & -1 & 2 \\ -1 & -3 & -1 & 1 \end{pmatrix}$ with respect to the standard basis of $\mathbb{R}^4$ and the standard basis of $\mathbb{R}_3[x]$.

Find the image of $\boldsymbol{v} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ -2 \end{pmatrix}$.

$[T(\boldsymbol{v})]_{\mathcal{S}_{\mathbb{R}_3[x]}} = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 3 & 1 & 0 \\ 2 & 3 & -1 & 2 \\ -1 & -3 & -1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \\ 7 \\ -12 \end{pmatrix}$. Therefore $T(\boldsymbol{v}) = 5 + 10x + 7x^2 - 12x^3$.

Method 6.3.3. Given the matrix A of a linear transformation $T : V \to V$ with respect to a basis $\mathcal{B}$ for V and given a second basis $\mathcal{B}'$ for V compute the matrix of T with respect to $\mathcal{B}'$.

If $P = P_{\mathcal{B}' \to \mathcal{B}}$ is the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$ then the matrix A' of T with respect to $\mathcal{B}'$ is $P^{-1}AP$.

Example 6.3.9. Let $T : \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ be a linear transformation and have matrix $A = \begin{pmatrix} 3 & -2 & 1 \\ 1 & 0 & -4 \\ 4 & 5 & 6 \end{pmatrix}$ with respect to the standard basis $(1, x, x^2)$ of $\mathbb{R}_2[x]$. Find the matrix A' of T with respect to the basis

$$\mathcal{B} = (1 + x + 2x^2, 2 + 3x + 2x^2, 2 + x + 5x^2).$$

The change of basis matrix $P = P_{\mathcal{B} \to \mathcal{S}}$ is easy to write down: $P_{\mathcal{B} \to \mathcal{S}} = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 2 & 5 \end{pmatrix}$.

The matrix P^{-1} is obtained as the inverse of P using the inversion algorithm. In this case we obtain $P^{-1} = \begin{pmatrix} -13 & 6 & 4 \\ 3 & -1 & -1 \\ 4 & -2 & -1 \end{pmatrix}$.

Now

$$A' = \begin{pmatrix} -13 & 6 & 4 \\ 3 & -1 & -1 \\ 4 & -2 & -1 \end{pmatrix} \begin{pmatrix} 3 & -2 & 1 \\ 1 & 0 & -4 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 1 & 3 & 1 \\ 2 & 2 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 78 & -53 \\ -5 & -23 & -52 \\ 5 & -15 & 29 \end{pmatrix}.$$

Example 6.3.10. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a **linear transformation** and have **matrix** $A = \begin{pmatrix} 7 & 4 & 5 \\ -1 & 2 & -1 \\ -2 & -2 & 0 \end{pmatrix}$ with respect to the **standard basis** of $\mathbb{R}^3$. Compute the **matrix** A' of T with respect to the **basis** $\mathcal{B} = \left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ -1 \end{pmatrix} \right)$.

If $P = P_{\mathcal{B} \to \mathcal{S}} = \begin{pmatrix} -1 & -1 & 3 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix}$ then by the **inversion algorithm** we obtain $P_{\mathcal{S} \to \mathcal{B}} = P^{-1} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

Then

$$A' = P^{-1}AP = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 7 & 4 & 5 \\ -1 & 2 & -1 \\ -2 & -2 & 0 \end{pmatrix} \begin{pmatrix} -1 & -1 & 3 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Method 6.3.4. Given the **matrix** A of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W and the **coordinate vector** of $\boldsymbol{v} \in V$ determine if $\boldsymbol{v}$ is in the **kernel** of T.

Multiply A by the **coordinate vector** $[\boldsymbol{v}]_{\mathcal{B}}$ to see if this is in the **null space** of A.

Example 6.3.11. Let $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}_2[x]$ have the **matrix** $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & 2 & 5 & 0 \\ 2 & 3 & 0 & 5 \end{pmatrix}$ with respect to the **standard basis** of $M_{2\times 2}(\mathbb{R})$ and the **standard basis** of $\mathbb{R}_2[x]$.

Determine if the following are in the **kernel** of T:

a) $\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$;

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

a) The **coordinate vector** is $\begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}$. The product of A with this coordinate vector is $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & 2 & 5 & 0 \\ 2 & 3 & 0 & 5 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. So the **coordinate vector** is in the **null space** of A and the vector $\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \in M_{2\times 2}(\mathbb{R})$ is in the **kernel** of T.

b) The **coordinate vector** is $\begin{pmatrix} 2 \\ 2 \\ -1 \\ -3 \end{pmatrix}$.

The product of A with this coordinate vector is $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & 2 & 5 & 0 \\ 2 & 3 & 0 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ -1 \\ -3 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \\ -5 \end{pmatrix}$.

Therefore, $\begin{pmatrix} 2 & -1 \\ 2 & -3 \end{pmatrix}$ is not in the **kernel of T**.

Method 6.3.5. Given the **matrix** of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W and the **coordinate vector** of $\boldsymbol{w} \in W$ determine if $\boldsymbol{w}$ is in the **range** of T.

Use **Theorem** (2.3.1) to determine if the **coordinate vector** $[\boldsymbol{w}]_{\mathcal{B}_W}$ is in the **span** of the columns of A (equivalently, in the **column space** of A). To do this, augment A by the **coordinate vector** of $\boldsymbol{w}$ and use **Gaussian elimination** to obtain an **echelon form** and determine if the last column is a **pivot column**.

Example 6.3.12. Let $T : \mathbb{R}_2[x] \to M_{2\times 2}(\mathbb{R})$ have matrix $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 3 & 0 \\ 3 & 4 & 5 \end{pmatrix}$ with respect to the **standard basis** of $\mathbb{R}_2[x]$ and the **standard basis** of $M_{2\times 2}(\mathbb{R})$.

Determine if the following vectors are in the **range** of T.

a) $\begin{pmatrix} 3 & 4 \\ 7 & 7 \end{pmatrix}$

b) $\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$.

We do both at the same time by adjoining both **coordinate vectors**. The matrix we obtain is $\begin{pmatrix} 1 & 2 & 1 & | & 3 & 1 \\ 2 & 5 & 1 & | & 7 & 3 \\ 1 & 3 & 0 & | & 4 & 2 \\ 3 & 4 & 5 & | & 7 & 2 \end{pmatrix}$.

The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 3 & | & 1 & 0 \\ 0 & 1 & -1 & | & 1 & 0 \\ 0 & 0 & 0 & | & 0 & 1 \\ 0 & 0 & 0 & | & 0 & 0 \end{pmatrix}$.

So, we conclude that $\begin{pmatrix} 3 & 4 \\ 7 & 7 \end{pmatrix} \in R(T)$ and $\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \notin R(T)$.

Method 6.3.6. Given the **matrix** of a **linear transformation** $T : V \to W$ with respect to **bases** $\mathcal{B}_V$ for V and $\mathcal{B}_W$ for W determine if T is **onto**, **one-to-one**, or an **isomorphism**.

i) Let $dim(V) = n, dim(W) = m$ then A is an $m \times n$ matrix. T is **onto** if and only if $col(A) = \mathbb{R}^m$. If $m > n$ then this is impossible and there is nothing to compute. If $m \leq n$ use **Gaussian elimination** to obtain an **echelon form** of A. Use this to determine if every row of A contains a **pivot position**. If so, then $col(A) = \mathbb{R}^m$ and T is **onto**, otherwise it is not.

ii) T is **one-to-one** if and only if $null(A) = \{\mathbf{0}_n\}$ if and only if every column of A is a **pivot column**. If $n > m$ then this is not possible and there is nothing to compute. If $n \leq m$ then use **Gaussian elimination** to obtain an **echelon form** of A. If every column of A is a **pivot column** then T is **one-to-one**, otherwise it is not.

iii) T is an **isomorphism** if and only if A is **invertible**. If $m \neq n$ this is not possible (only square matrices can be invertible). If $m = n$ use **Gaussian elimination** to obtain an **echelon form** of A. If every row contains a **pivot position**, equivalently, every column is a **pivot column**, then T is an **isomophism**, otherwise it is not.

Example 6.3.13. Let $T : \mathbb{R}_2[x] \to M_{2\times 2}(\mathbb{R})$ have matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 5 & 1 \\ 1 & 2 & 0 \\ 4 & 3 & 5 \end{pmatrix}$ with respect to the **standard basis** of $\mathbb{R}_2[x]$ and the **standard basis** of $M_{2\times 2}(\mathbb{R})$. Determine if T is **one-to-one** or **onto**.

Since there are fewer columns than rows T cannot be onto. Using **Gaussian elimination** we obtain the **reduced echelon form**: $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Since not every column is a **pivot columns**, this transformation is not **one-to-one**.

Example 6.3.14. Let $T : M_{2\times 2} \to \mathbb{R}^4$ have matrix $A = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 2 & 5 & 3 & 1 \\ 2 & 4 & 1 & 3 \\ 1 & 3 & 0 & 1 \end{pmatrix}$ with respect to the **standard matrix** of $M_{2\times 2}(\mathbb{R})$ and the **standard basis** of $\mathbb{R}^4$. Determine if T is an **isomorphism**.

The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. The **linear transformation** T is an **isomorphism**.

Exercises

In exercises 1 - 7 write down the **matrix** of the given **linear transformation** with respect to the standard bases of the **domain and codomain**. See **Method** (6.3.1).

1. $T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a - 2b + 3c + d \\ 2a - 3b + 5c + 3d \\ -2a + 4b - 6c - 2d \\ 4a - 5b + 9c + 6d \end{pmatrix}$.

2. $T(a + bx + cx^2) = \begin{pmatrix} 2a + 5b + c & -a - 3b - 2c \\ 3a + 7b + c & -4a - 9b - c \end{pmatrix}$

3. $T(a+bx+cx^2) = \begin{pmatrix} a+2b+c \\ 4a+7b+5c \\ 3a+5b+5c \end{pmatrix}$

4. $T(1) = \begin{pmatrix} 1 & 3 \\ 2 & 3 \end{pmatrix}, T(x) = \begin{pmatrix} 2 & 6 \\ 5 & 8 \end{pmatrix}, T(x^2) = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}, T(x^3) = \begin{pmatrix} 1 & 4 \\ 0 & -1 \end{pmatrix}$

5. $T(1) = 1+4x-2x^2, T(x) = 2+7x+x^2, T(x^2) = 1+5x-6x^2$

6. $T(a+bx) = (11a-6b)+(18a-10b)x$

7. $T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x-10y+8z \\ 4x-18y+13z \\ 4x-20y+15z \end{pmatrix}$

In exercises 8 - 11 let $T : \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ have **matrix** $\begin{pmatrix} 1 & -2 & 1 \\ -2 & 3 & 1 \\ 1 & -3 & 4 \end{pmatrix}$ with respect to the basis $\mathcal{B} = \{1+x, x+2x^2, 2+x-x^2\}$.

8. Compute the **image** of $2-3x+4x^2$. See **Method** (6.3.2).

9. Find the **matrix** of T with respect to the **standard basis**, $(1, x, x^2)$, of $\mathbb{R}_2[x]$. See **Method** (6.3.3).

10. If $[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix}$ determine if $\boldsymbol{v} \in Ker(T)$. See **Method** (6.3.4).

11. If $[\boldsymbol{w}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}$ is $\boldsymbol{w} \in R(T)$. If so, find a vector $\boldsymbol{v} \in V$ such that $T(\boldsymbol{v}) = \boldsymbol{w}$. See **Method** (6.3.5).

12. Find the **matrix** of the **linear transformation** of exercise 6 with respect to the **basis** $\mathcal{B}' = (1+2x, 2+3x)$. See **Method** (6.3.3).

13. Find the **matrix** of the **linear transformation** of exercise 7 with respect to the **basis** $\mathcal{B}' = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right)$. See **Method** (6.3.3).

In exercises 14-23 see **Method** (6.3.6).

14. Determine if the **linear transformation** of exercise 1 is **one-to-one** or **onto**.

15. Determine if the **linear transformation** of exercise 2 is **one-to-one** or **onto**.

16. Determine if the **linear transformation** of exercise 3 is **one-to-one** or **onto**.

17. Determine if the **linear transformation** of exercise 4 is **one-to-one** or **onto**.

18. Determine if the **linear transformation** of exercise 5 is **one-to-one** or **onto**.

19. Determine if the **linear transformation** of exercise 6 is **one-to-one** or **onto**.

20. Determine if the **linear transformation** of exercise 7 is **one-to-one** or **onto**.

21. Determine if the **linear transformation** of exercise 1 is an **isomorphism**.

22. Determine if the **linear transformation** of exercise 3 is an **isomorphism**..

23. Determine if the **linear transformation** of exercise 4 is an **isomorphism**.

In exercises 24 - 27 answer true or false and give an explanation.

24. If $T : V \to W$ is a **linear transformation** and is **one-to-one** then $dim(V) > dim(W)$.

25. If $T : V \to V$ is a **linear transformation** with matrix A with respect to the **basis** $\mathcal{B}$ of V and $\boldsymbol{v} \in Ker(T)$ then $[\boldsymbol{v}]_{\mathcal{B}} \in null(A)$.

26. If $T : V \to V$ is a **linear transformation** with matrix A with respect to the **basis** $\mathcal{B}$ of V and $[\boldsymbol{v}]_{\mathcal{B}} \in col(A)$ then $\boldsymbol{v} \in R(T)$.

27. Assume V and W are vector spaces, V is **4-dimensional** with **basis** $\mathcal{B}_V$, and W is 5-dimensional with basis $\mathcal{B}_W$, $T : V \to W$ is a **linear transformation**, and A is the **matrix of T** with respect to $(\mathcal{B}_V, \mathcal{B}_W)$. Then $A$is a 4×5 matrix.

Challenge Exercises (Problems)

1. Let U, V, W be **finite dimensional vector spaces** with **bases** $\mathcal{B}_U, \mathcal{B}_V, \mathcal{B}_W$, respectively. Assume that $S : U \to V$ is a **linear transformation** and $T : V \to W$ is a linear tranformation. Prove that if A is the **matrix** of S with respect to $\mathcal{B}_U$ and $\mathcal{B}_V$ and B is the matrix of T with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$ then BA is the matrix of $T \circ S : U \to W$ with respect to $\mathcal{B}_U$ and $\mathcal{B}_W$.

2. Let V and W be **vector spaces** of **dimension** n and $\mathcal{B}_V$ a **basis** of V and $\mathcal{B}_W$ a basis of W. Assume that $T : V \to W$ is an **isomorphism** and A is the **matrix** of T with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$. Prove that A is **invertible** and that A^{-1} is the **matrix** of $T^{-1} : W \to V$ with respect to $\mathcal{B}_W$ and $\mathcal{B}_V$.

3. Let V be a **vector space** of **dimension** n and W a vector space of dimension m, $\mathcal{B}_V$ a **basis** of V and $\mathcal{B}_W$ a basis of W. Assume that $S : V \to W$ and $T : V \to W$ are linear transformations. Prove that if A is the **matrix** of S with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$ and B is the matrix of T with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$ then $A + B$ is the matrix of $S + T : V \to W$ with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$.

4. Suppose V is a **two dimensional vector space** with **basis** $\mathcal{B}$ and $T : V \to V$ is a **linear transformation** with **matrix** A with respect to $\mathcal{B}$. Suppose $\boldsymbol{v}_1, \boldsymbol{v}_2$ are non-zero vectors which satisfy $T(\boldsymbol{v}_1) = \boldsymbol{v}_1, T(\boldsymbol{v}_2) = 2\boldsymbol{v}_2$.

a) Prove that $\mathcal{B}' = (\boldsymbol{v}_1, \boldsymbol{v}_2)$ is **linearly independent** and therefore a **basis** of V.

b) Let $P = P_{\mathcal{B}\to\mathcal{B}'}$ be the **change of basis matrix** from $\mathcal{B}$ to $\mathcal{B}'$. Prove that $PAP^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

5. Assume that A and B are **similar** $n \times n$ matrices and $B = P^{-1}AP$.

a) For a postive integer m prove that $B^m = P^{-1}A^mP$.

b) Assume that A is **invertible**. Prove that B is **invertible** and that $B^{-1} = P^{-1}A^{-1}P$.

Quiz Solutions

1. $\begin{pmatrix} 2 & 4 & 7 & -5 \\ 3 & -6 & 9 & 2 \\ -4 & 1 & 8 & -3 \end{pmatrix} \begin{pmatrix} 3 \\ -5 \\ 7 \\ 2 \end{pmatrix} = \begin{pmatrix} 25 \\ 106 \\ 33 \end{pmatrix}$

Not right, see **Method** (3.1.1).

2. The matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 0 & 19 & 8 \\ 0 & 1 & 0 & -3 & -1 \\ 0 & 0 & 1 & -4 & -3 \end{pmatrix}$.

Then $null[\begin{pmatrix} 1 & 3 & 2 & 2 & -1 \\ 2 & 5 & 5 & 3 & -4 \\ 2 & 7 & 4 & 1 & -3 \end{pmatrix}] = Span\left(\begin{pmatrix} -19 \\ 3 \\ 4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -8 \\ 1 \\ 3 \\ 0 \\ 1 \end{pmatrix}\right)$.

Not right, see **Method** (3.2.2).

3. This is the same as asking if the vector is in the **span** of the columns. The augmented matrix $\begin{pmatrix} 1 & 1 & 1 & 1 & | & 1 \\ 2 & 3 & 4 & 4 & | & 8 \\ 4 & 3 & 3 & 5 & | & 7 \\ 3 & 2 & 1 & 1 & | & -3 \\ 7 & 5 & 4 & 6 & | & 4 \end{pmatrix}$ has **reduced echelon form**

$\begin{pmatrix} 1 & 0 & 0 & 2 & | & 4 \\ 0 & 1 & 0 & -4 & | & -12 \\ 0 & 0 & 1 & 3 & | & 9 \\ 0 & 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$

Yes, $\begin{pmatrix} 1 \\ 8 \\ 7 \\ -3 \\ 4 \end{pmatrix} \in col(A)$.

Not right, see **Method** (2.2.1).

Chapter 7

Eigenvalues and Eigenvectors

7.1. Introduction to Eigenvalues and Eigenvectors

In this section we introduce the concept of an eigenvalue and eigenvector of a square matrix.

Am I Ready for This Material

The main concepts that will be used in this section are

null space of a matrix

diagonal matrix

basis of a vector space

change of basis matrix

determinant of a 2 x 2 matrix

determinant of a 3 x 3 matrix

determinant of a 4 x 4 matrix

determinant of an n x n matrix

Quiz

In 1 - 3 compute the **determinant of an n x n matrix**

1. $\begin{pmatrix} 2 & 4 & 7 \\ 2 & 1 & -3 \\ 5 & 2 & 6 \end{pmatrix}$

2. $\begin{pmatrix} 2-\lambda & 7 \\ -2 & -1-\lambda \end{pmatrix}$

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

4. Factor the following polynomials:

a) $\lambda^2 + 3\lambda - 10$

b) $\lambda^3 - 7\lambda + 6$

5. Find a **basis** for the **null space** of the matrix

$$A = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 3 & 1 & 1 \\ 4 & 7 & 3 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

Quiz Solutions

New Concepts

This section introduces several important concepts which are extremely important and applied throughout mathematics, science, and engineering. These are:

eigenvector of a (square) matrix

eigenvalue of a (square) matrix

characteristic polynomial of a (square) matrix

algebraic multiplicity of a real eigenvalue

eigenspace of a (square) matrix

geometric multiplicity of a real eigenvalue

Theory (Why It Works)

Let A be an $n \times n$ matrix. It is clearly easier to understand the action of A as a **matrix transformation** when A is a **diagonal matrix**:

$$A = diag(a_1, a_2, \dots, a_n) = \begin{pmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & a_n \end{pmatrix}.$$

Specifically, when we apply A to the **standard basis vector** $\boldsymbol{e}_i$ we get a multiple of $\boldsymbol{e}_i$ and so A is a combination of expansions and/or contractions in the directions of the standard basis vectors: $A\boldsymbol{e}_i = a_i\boldsymbol{e}_i$.

It is our purpose in this section to determine for an arbitrary $n \times n$ matrix A whether there are non-zero vectors on which A acts in this way. That is, we want to determine if there vectors $\boldsymbol{v} \neq \boldsymbol{0}_n$ such that $A\boldsymbol{v}$ is a multiple of $\boldsymbol{v}$. We will especially be interested in whether we can find a **basis** for $\mathbb{R}^n$ consisting of such vectors and this is subject of the next section. This concept is one of the most important in linear algebra and is extensively applied in chemistry physics and engineering. This idea is the subject of the following definition:

Definition 7.1. Let A be an $n \times n$ matrix. A nonzero **n-vector** $\boldsymbol{v}$ is an ***eigenvector*** with ***eigenvalue*** λ if $A\boldsymbol{v} = \lambda\boldsymbol{v}$. This means that when $\boldsymbol{v}$ is multiplied on the left by the matrix A then the result is a scalar multiple of $\boldsymbol{v}$.

Remark 7.1. Assume that $\boldsymbol{v}$ is an **eigenvector** of the $n \times n$ matrix A with **eigenvalue** λ and α is a real number. Then $\boldsymbol{v}$ is an eigenvector of the matrix $A - \alpha I_n$ with eigenvalue $\lambda - \alpha$.

Example 7.1.1. 1. Let $A = \begin{pmatrix} 5 & -2 \\ -3 & 10 \end{pmatrix}, \boldsymbol{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$. Then

$$A\boldsymbol{v}_1 = \begin{pmatrix} 5 & -2 \\ -3 & 10 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \end{pmatrix} = 4\boldsymbol{v}_1,$$

$$A\boldsymbol{v}_2 = \begin{pmatrix} 5 & -2 \\ -3 & 10 \end{pmatrix}\begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} -11 \\ 33 \end{pmatrix} = 11\boldsymbol{v}_2.$$

Thus, $\boldsymbol{v}_1$ is an **eigenvector with eigenvalue** 4 and $\boldsymbol{v}_2$ is an eigenvector with eigenvalue 11.

Example 7.1.2. Let $A = \begin{pmatrix} 2 & -1 & -2 \\ -1 & 2 & 2 \\ 2 & -2 & -3 \end{pmatrix}, \boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$.

Then

$$A\boldsymbol{v}_1 = \boldsymbol{v}_1, A\boldsymbol{v}_2 = \boldsymbol{v}_2, A\boldsymbol{v}_3 = -\boldsymbol{v}_3$$

(check these).

Therefore $\boldsymbol{v}_1, \boldsymbol{v}_2$ are **eigenvectors with eigenvalue 1** and $\boldsymbol{v}_3$ is an eigenvector with eigenvalue -1.

Example 7.1.3. Let $A = \begin{pmatrix} -6 & 6 & -6 \\ -8 & 8 & -6 \\ -5 & 5 & -3 \end{pmatrix}$ and $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $\boldsymbol{v}_3 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$. Then

$$A\boldsymbol{v}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, A\boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} = 2\boldsymbol{v}_2, A\boldsymbol{v}_3 = \begin{pmatrix} -6 \\ -6 \\ -3 \end{pmatrix} = -3\boldsymbol{v}_3.$$

Therefore $\boldsymbol{v}_1$ is an **eigenvector with eigenvalue 0**, $\boldsymbol{v}_2$ is an eigenvector with eigenvalue 2, and $\boldsymbol{v}_3$ is an eigenvector with eigenvalue -3.

Determining the eigenvalues of a matrix

Suppose A is an $n \times n$ matrix and $\boldsymbol{v}$ is an **eigenvector with eigenvalue λ.** This means that $A\boldsymbol{v} = \lambda\boldsymbol{v} = (\lambda I_n)\boldsymbol{v}$. From this we conclude that

$$A\boldsymbol{v} - (\lambda I_n)\boldsymbol{v} = [A - (\lambda I_n)]\boldsymbol{v} = \mathbf{0}_n.$$

Since $\boldsymbol{v}$ is nonzero, it follows that the **null space** of $A - \lambda I_n$ is not trivial.

Recall, the **null space** of an $m \times n$ matrix B is the collection of vectors $\boldsymbol{w}$ such that $B\boldsymbol{w} = \mathbf{0}_m$. The **null space** of a square matrix is nontrivial if and only if the matrix is **non-invertible** which occurs if and only if the **determinant** of B is zero.

This gives us a criteria for a (real) number λ to be an **eigenvalue**:

Theorem 7.1.1. *The real number λ is an* **eigenvalue** *of the $n \times n$ matrix A if and only if $det(A - \lambda I_n) = 0$.*

Proof. We have already shown that if λ is an **eigenvalue** of A then $det(A - \lambda I_n) = 0$. On the other hand, suppose λ is a real number and $\det(A - \lambda I_n) = 0$. Then the matrix $A - \lambda I_n$ is **non-invertible** and so its **null space** is nontrivial. If $\boldsymbol{v} \neq \mathbf{0}_n$ in the **null space** of $A - \lambda I_n$ we then have

$$(A - \lambda I_n)\boldsymbol{v} = A\boldsymbol{v} - \lambda I_n\boldsymbol{v} =$$

$$A\boldsymbol{v} - \lambda\boldsymbol{v} = \mathbf{0}_n \tag{7.1}$$

From (7.1) we conclude that $A\boldsymbol{v} = \lambda\boldsymbol{v}$ and therefore $\boldsymbol{v}$ is an **eigenvector with eigenvalue** λ. □

Definition 7.2. Let A be an $n \times n$ matrix. The polynomial $\chi_A(\lambda) = det(A - \lambda I_n)$ is called the ***characteristic polynomial of*** A. The equation $\chi_A(\lambda) = 0$ is called the ***characteristic equation***. If α is a real number and a root of $\chi_A(\lambda)$ and $(\lambda - \alpha)^e$ divides $\chi_A(\lambda)$ but $(\lambda - \alpha)^{e+1}$ does not divide $\chi_A(\lambda)$ then e is said to be the ***algebraic multiplicity of the eigenvalue*** α. This is the number of times that α is a root of the polynomial $\chi_A(\lambda)$.

Example 7.1.4. Find the **eigenvalues** of $A = \begin{pmatrix} 4 & -3 & -1 \\ -2 & 9 & 2 \\ -1 & 3 & 4 \end{pmatrix}$.

We form $A - \lambda I_3$ and compute its **determinant**. $A - \lambda I_3 = \begin{pmatrix} 4-\lambda & -3 & -1 \\ -2 & 9-\lambda & 2 \\ -1 & 3 & 4-\lambda \end{pmatrix}$.

The **determinant** of $A - \lambda I_3$ is

$$(4-\lambda)(9-\lambda)(4-\lambda) + 6 + 6 - (9-\lambda) - 6(4-\lambda) - 6(4-\lambda) =$$

$$-\lambda^3 + 17\lambda^2 - 88\lambda + 144 - 9 + \lambda - 24 + 6\lambda - 24 + 6\lambda = -\lambda^3 + 17\lambda^2 - 75\lambda + 99.$$

This factors as $-(\lambda - 3)^2(\lambda - 11)$.

Thus $\lambda = 3$ is an **eigenvalue** and it occurs twice as a root of $det(A - \lambda I_3) = 0$. Therefore the **eigenvalue** 3 occurs with **algebraic multiplicity** two. On the other hand, 11 occurs as a root of the **characteristic polynomial** exactly once and so has **algebraic multiplicity** one.

Definition 7.3. Let α be an **eigenvalue** of the $n \times n$ matrix A. The collection of vectors $\{\boldsymbol{v} \in \mathbb{R}^n : A\boldsymbol{v} = \alpha\boldsymbol{v}\}$ is called the α*–**eigenspace of*** A and is denoted by $E_\alpha(A) = E_\alpha$.

Remark 7.2. Let A be an $n \times n$ matrix and α an **eigenvalue** of A. Then the zero vector $\mathbf{0}_n$ belongs to the α-eigenspace but it is not an **eigenvector** since the definition requires an eigenvector to be a non-zero vector. Every other vector in E_α is an **eigenvector with eigenvalue α**.

Theorem 7.1.2. *Let A be an $n \times n$ matrix with* **eigenvalue** *α. Then E_α is the* **null space** *of $A - \alpha I_n$. In particular, E_α is a* **subspace** *of $\mathbb{R}^n$.*

Proof. Suppose that $\boldsymbol{v} \in null(A - \alpha I_n)$. Then $(A - \alpha I_n)\boldsymbol{v} = \mathbf{0}_n$. On the other hand, $(A - \alpha I_n)\boldsymbol{v} = A\boldsymbol{v} - \alpha\boldsymbol{v}$ and consequently, $A\boldsymbol{v} = \alpha\boldsymbol{v}$ and so $\boldsymbol{v} \in E_\alpha$. Conversely, if $\boldsymbol{v} \in E_\alpha$ then $A\boldsymbol{v} = \alpha\boldsymbol{v} = (\alpha I_n)\boldsymbol{v}$ and therefore $(A - \alpha I_n)\boldsymbol{v} = \mathbf{0}_n$ which implies that $\boldsymbol{v} \in null((A - \alpha I_n)$. $\square$

Definition 7.4. Let A be an $n \times n$ matrix and $\alpha \in \mathbb{R}$ an eigenvalue of A. The **dimension** of E_α is the ***geometric multiplicity*** of α.

Remark 7.3. The **geometric multiplicity** of an **eigenvalue** α is always less than or equal to the **algebraic multiplicity** of α. Moreover, the algebraic and geometric multiplicity may or may not be equal.

Example 7.1.5. Find the **eigenspaces** for the matrix A of Example (7.1.4)

We have found that the **eigenvalues** are 3 with **algebraic multiplicity** 2 and 11 with algebraic multiplicity 1. We substract these from the diagonal and, in the respective cases, find the **null space** of the resulting matrix.

$\lambda = 3 : A - 3I_3 = \begin{pmatrix} 1 & -3 & -1 \\ -2 & 6 & 3 \\ -1 & 3 & 1 \end{pmatrix}$. We apply **Gaussian elimination** to this matrix to obtain the **reduced echelon form** which is $\begin{pmatrix} 1 & -3 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

This matrix has one **leading variable** and two **free variables**. Therefore E_3 is **spanned** by two vectors, has **dimension** two and the **geometric multiplicity** of the **eigenvalue** 3 is also two. We proceed to systematically find a pair of such vectors, that is a **basis** , in the usual way.

The corresponding system (equation) is $x - 3y - z = 0$. Choosing parameters $y = s, z = t$ we get the solution

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3s+t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Thus, $E_3 = Span\left(\begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\right)$.

$\lambda = 11 : A - 11I_3 = \begin{pmatrix} -7 & -3 & -1 \\ -2 & -2 & 2 \\ -1 & 3 & -7 \end{pmatrix}$.

Applying **Gaussian elimination** we obtain the **reduced echelon form** of the matrix $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$.

We have two **leading variables** and one **free variable**. Therefore a single vector will **span** the **eigenspace** for the eigenvalue 11. We proceed to determine such a vector. The **homogeneous linear system** with **coefficient matrix** $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$ is

$$\begin{array}{ccccccc} x & & & + & z & = & 0 \\ & & y & - & 2z & = & 0 \end{array}.$$

Setting $z = t$ we get the solutions $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -t \\ 2t \\ t \end{pmatrix} = t\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$.

Therefore the vector $\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$ **spans** the 11-eigenspace, E_{11}.

Eigenvectors and Eigenvalues of a Linear Operator

Since **linear transformations** and matrices are in a very real sense two ways of looking at the same object it should not be surprising that all of the definitions we have made for square matrices make sense for a **linear operator** $T : V \to V$. This is the thrust of the following definition:

Definition 7.5. Let $T : V \to V$ be a **linear operator**. A nonzero vector $\boldsymbol{v}$ in V is an ***eigenvector*** with ***eigenvalue*** λ if $T(\boldsymbol{v}) = \lambda\boldsymbol{v}$.

The **eigenvalues** for a **linear operator** $T : V \to V$ can be found by choosing any **basis** $\mathcal{B}$ for V and computing the **matrix** of T with respect to $\mathcal{B}$ and using the methods of this section to find the **eigenvalues** of A. We state this as a theorem:

Theorem 7.1.3. *Let V be an **finite dimensional vector space** with **basis** $\mathcal{B}$ and $T : V \to V$ a **linear operator**. Assume that the **matrix** of T with respect to $\mathcal{B}$ is A. Then a vector $\boldsymbol{v} \in V$ is an **eigenvector** of T with **eigenvalue** α if and only if the **coordinate vector** of $\boldsymbol{v}$ with respect to $\mathcal{B}$ is an **eigenvector** of A with **eigenvalue** α.*

Proof. By the definition of A, $[T(\boldsymbol{v})]_{\mathcal{B}} = A[\boldsymbol{v}]_{\mathcal{B}}$. If $T(\boldsymbol{v}) = \alpha\boldsymbol{v}$ then $[T(\boldsymbol{v})]_{\mathcal{B}} = [\alpha\boldsymbol{v}]_{\mathcal{B}} = \alpha[\boldsymbol{v}]_{\mathcal{B}}$. Therefore $A[\boldsymbol{v}]_{\mathcal{B}} = \alpha[\boldsymbol{v}]_{\mathcal{B}}$ and $[\boldsymbol{v}]_{\mathcal{B}}$ is an **eigenvector** of A with **eigenvalue** α.

On the other hand, if $[\boldsymbol{v}]_{\mathcal{B}}$ is an **eigenvector** of A with **eigenvalue** α then $[T(\boldsymbol{v})]_{\mathcal{B}} = A[\boldsymbol{v}]_{\mathcal{B}} = \alpha[\boldsymbol{v}]_{\mathcal{B}}$ and therefore $T(\boldsymbol{v}) = \alpha\boldsymbol{v}$ and $\boldsymbol{v}$ is an eigenvector of T with eigenvalue α. □

Example 7.1.6. Let $T : \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ be the **linear operator** such that

$$T(1) = 2 + x + x^2$$
$$T(x) = 1 + 2x + x^2$$
$$T(x^2) = 1 + x + 2x^2$$

Let $\mathcal{S} = (1, x, x^2)$, the **standard basis** of $\mathbb{R}_2[x]$. Set $A = \mathcal{M}_T(\mathcal{S}, \mathcal{S})$, the **matrix** of T with respect to $\mathcal{S}$. Then

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

We compute the **characteristic polynomial** of A in order to find the **eigenvalues** and their respective **algebraic multiplicities**:

$$A - \lambda I_3 = \begin{pmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{pmatrix}$$

Then $det(A - \lambda I_3) = (2-\lambda)^3 + 1 + 1 - (2-\lambda) - (2-\lambda) - (2-\lambda) =$

$$-\lambda^3 + 6\lambda^2 - 9\lambda + 4 = -(\lambda - 1)^2(\lambda - 4)$$

Thus, 1 is an **eigenvalue** with **algebraic multiplicity** two and 4 is an eigenvalue with algebraic multiplicity one. We proceed to find the respective **eigenspaces**.

The **eigenspace** for the **eigenvalue** 1:

$A - I_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. The **homogeneous linear system** with **coefficient matrix** $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is

$$x + y + z = 0$$

There is one **leading variable** (x) and two **free variables** (y, z). We set $y = s, z = t$ and solve for all the variables in terms of s and t. We get

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -s-t \\ s \\ t \end{pmatrix} =$$

$$s\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

It therefore follows that $\left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right)$ is a **basis** for $E_1(A)$, the **1-eigenspace** of A. We conclude that the **geometric multiplicity** of 1 is two.

The **eigenspace** for the **eigenvalue** 4:

$A - 4I_3 = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$.

The **homogeneous linear system** with **coefficient matrix** $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$ is

$$\begin{array}{ccccccc} x & & & - & z & = & 0 \\ & & y & - & z & = & 0 \end{array}$$

There are two **leading variables** (x, y) and one **free variables** (z). We set $z = t$ and solve for all the variables in terms of t. We get

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} t \\ t \\ t \end{pmatrix} = t \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Thus, $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ **spans** the **eigenspace** $E_4(A)$.

We now return to the operator T. It follows from **Theorem** (7.1.3) that the **eigenvalues** of T are 1 with algebraic multiplicity two and 4 with algebraic multiplicity one. Moreover, $(-1 + x, -1 + x^2)$ is a **basis** for the $E_1(T)$ and $1 + x + x^2$ **spans** $E_4(T)$. It is straightforward to check that $\mathcal{B} = (-1 + x, -1 + x^2, 1 + x + x^2)$ is a **basis** for $\mathbb{R}_2[x]$. Since

$$T(-1 + x) = -1 + x$$
$$T(-x + x^2) = -x + x^2$$
$$T(1 + x + x^2) = 4 + 4x + 4x^2 = 4(1 + x + x^2)$$

it follows that $\mathcal{M}_T(\mathcal{B}, \mathcal{B}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$.

Remark 7.4. All the definitions and theorems proved in this section apply to square matrices over fields other than $\mathbb{R}$ (the real numbers), in particular they apply to $\mathbb{C}$, the **field of complex numbers**.

What You Can Now Do

1. Compute the **characteristic polynomial** of a square matrix.

2. Determine the **eigenvalues** of a square matrix as well as the **algebraic multiplicity** of each distinct eigenvalue.

3. For each **eigenvalue** α of a square matrix A find a **basis** for the **eigenspace** E_α and the **geometric multiplicity** of α.

Method (How To Do It)

Method 7.1.1. Compute the **characteristic polynomial** of a square matrix.

If the matrix is A, subtract λ from the diagonal elements (this is the matrix $A - \lambda I_n$.) Then take the **determinant**. The resulting polynomial in λ is the **characteristic polynomial**.

Example 7.1.7. Find the **characteristic polynomial** of the matrix $A = \begin{pmatrix} 5 & -4 \\ 3 & -2 \end{pmatrix}$.

We form the matrix $A - \lambda I_2 = \begin{pmatrix} 5-\lambda & -4 \\ 3 & -2-\lambda \end{pmatrix}$ and take the **determinant** :
$(5-\lambda)(-2-\lambda) + 12 = -10 - 3\lambda + \lambda^2 + 12 = \lambda^2 - 3\lambda + 2$.

Example 7.1.8. Find the **characteristic polynomial** of the matrix $A = \begin{pmatrix} 2 & 3 & -4 \\ 0 & 1 & -2 \\ 0 & 0 & 3 \end{pmatrix}$.

We form the matrix $A - \lambda I_3 = \begin{pmatrix} 2-\lambda & 3 & -4 \\ 0 & 1-\lambda & -2 \\ 0 & 0 & 3-\lambda \end{pmatrix}$ and take the **determinant** :
$(2-\lambda)(1-\lambda)(-3-\lambda)$.

Remark 7.5. For a **triangular matrix** with diagonal entries $d_1, d_2, \ldots, d_n$ the **characteristic polynomial** is $(d_1 - \lambda)(d_2 - \lambda)\ldots(d_n - \lambda)$. Don't bother multiplying out since in the end we will want to factor and find the roots of the characteristic polynomial.

Example 7.1.9. Find the **characteristic polynomial** of the matrix $A = \begin{pmatrix} -3 & 1 & 1 \\ -5 & 0 & 3 \\ -5 & -2 & 5 \end{pmatrix}$.

We form the matrix $A - \lambda I_3 = \begin{pmatrix} -3-\lambda & 1 & 1 \\ -5 & -\lambda & 3 \\ -5 & -2 & 5-\lambda \end{pmatrix}$ and take the **determinant** :

$$(-3-\lambda)(-\lambda)(5-\lambda) - 15 + 10 - 5\lambda + 6(-3-\lambda) + 5(5-\lambda) =$$

$$-\lambda^3 + 2\lambda^2 + 15\lambda - 15 + 10 - 5\lambda - 18 - 6\lambda + 25 - 5\lambda =$$

$$-\lambda^3 + 2\lambda^2 - \lambda + 2 = -(\lambda^3 - 2\lambda^2 + \lambda - 2) =$$

$$-(\lambda^2+1)(\lambda-2).$$

Remark 7.6. The outcome in Example (7.1.9) is intended to illustrate that the **characteristic polynomial** of a square matrix does not have to factor completely into (real) linear factors.

Method 7.1.2. Determine the **eigenvalues** of a square matrix as well as the **algebraic multiplicity** of each distinct eigenvalue.

Set the **characteristic polynomial** equal to zero to get the characteristic equation. Factor (if possible) and extract the real roots. More specifically, express the **characteristic polynomial** in the form $(-1)^n(\lambda-\alpha_1)^{e_1}(\lambda-\alpha_2)^{e_2}\dots(\lambda-\alpha_t)^{e_t}f(\lambda)$ where $f(\lambda)$ is a product of irreducible quadratics. The **eigenvalues** are α_1 with **algebraic multiplicity** e_1, α_2 with algebraic multiplicity $e_2, \dots, \alpha_t$ with algebraic multiplicity e_t.

Example 7.1.10. Find the real **eigenvalues** with their **algebraic multiplicities** for the matrix $\begin{pmatrix} 6 & 4 \\ -4 & -2 \end{pmatrix}$.

The **characteristic polynomial** is

$det \begin{pmatrix} 6-\lambda & 4 \\ -4 & -2-\lambda \end{pmatrix} = (6-\lambda)(-2-\lambda) + 16 = \lambda^2 - 4\lambda + 4 = (\lambda-2)^2.$

There is one real **eigenvalue**, 2, with **algebraic multiplicity** 2.

Example 7.1.11. Find the real **eigenvalues** with their **algebraic multiplicities** for the matrix $\begin{pmatrix} 3 & 5 \\ -1 & -3 \end{pmatrix}$.

The **characteristic polynomial** is

$$det \begin{pmatrix} 3-\lambda & 5 \\ -1 & -3-\lambda \end{pmatrix} = (3-\lambda)(-3-\lambda)+5 = \lambda^2 - 4 = (\lambda-2)(\lambda+2).$$

Now there are two distinct real **eigenvalues**, -2 and 2 each with **algebraic multiplicity** 1.

Example 7.1.12. Find the real **eigenvalues** with their **algebraic multiplicities** for the matrix $A = \begin{pmatrix} 5 & 5 \\ -2 & -1 \end{pmatrix}$.

The **characteristic polynomial** of this matrix is

$det \begin{pmatrix} 5-\lambda & 5 \\ -2 & -1-\lambda \end{pmatrix} = (5-\lambda)(-1-\lambda)+10 = \lambda^2 - 4\lambda + 5 = (\lambda-2)^2+1$. This quadratic has no real roots and so the matrix A has no real **eigenvalues**.

Example 7.1.13. Find the real **eigenvalues** with their **algebraic multiplicities** for the matrix $\begin{pmatrix} 5 & -4 & 8 \\ -4 & 1 & -4 \\ -4 & 2 & -5 \end{pmatrix}$.

The **characteristic polynomial** is $det \begin{pmatrix} 5-\lambda & -4 & 8 \\ -4 & 1-\lambda & -4 \\ -4 & 2 & -5-\lambda \end{pmatrix} =$

$(5-\lambda)(1-\lambda)(-5-\lambda) - 64 - 64 + 8(5-\lambda) - 16(-5-\lambda) + 32(1-\lambda) =$

$-\lambda^3 + \lambda^2 + 25\lambda - 25 - 64 - 64 + 40 - 8\lambda + 80 + 16\lambda + 32 - 32\lambda =$

$-\lambda^3 + \lambda^2 + \lambda - 1 = -(\lambda-1)^2(\lambda+1)$.

So there are two distinct real **eigenvalues** -1, 1. The **eigenvalue** -1 has **algebraic multiplicity** 1 and the eigenvalue 1 had algebraic multiplicity 2.

Example 7.1.14. Find the real **eigenvalues** with their **algebraic multiplicities** for the matrix $\begin{pmatrix} 0 & 1 & 0 \\ -2 & 3 & 0 \\ -2 & 1 & 2 \end{pmatrix}$.

The **characteristic polynomial** is

$det \begin{pmatrix} -\lambda & 1 & 0 \\ -2 & 3-\lambda & 0 \\ -2 & 1 & 2-\lambda \end{pmatrix} = (-\lambda)(3-\lambda)(2-\lambda) - (-2)(1)(2-\lambda) = -\lambda^3 + 5\lambda^2 - 8\lambda + 4.$

After fooling a little we find that 1 is a root and so $\lambda - 1$ a factor. The **characteristic polynomial** then becomes $-(\lambda-1)(\lambda^2 - 4\lambda + 4) = -(\lambda-1)(\lambda-2)^2$.

There are two distinct real **eigenvalues**, 1 and 2. The **eigenvalue** 1 has **algebraic multiplicity** 1 and the eigenvalue 2 has algebraic multiplicity 2.

Example 7.1.15. Find the real **eigenvalues** with their **algebraic multiplicities** for the matrix $\begin{pmatrix} 5 & 2 & -4 \\ 6 & 3 & -5 \\ 10 & 4 & -8 \end{pmatrix}$.

The **characteristic polynomial** is $det \begin{pmatrix} 5-\lambda & 2 & -4 \\ 6 & 3-\lambda & -5 \\ 10 & 4 & -8-\lambda \end{pmatrix} =$

$$(5-\lambda)(3-\lambda)(-8-\lambda) - 96 - 100 + 40(3-\lambda) - 12(-8-\lambda) + 20(5-\lambda) =$$

$$-\lambda^3 - 15\lambda + 40\lambda + 24\lambda - 120 - 96 - 100 + 120 - 40\lambda + 96 + 12\lambda + 100 - 20\lambda = -\lambda^3 + \lambda =$$

$$-\lambda(\lambda^2 - 1) = -(\lambda+1)\lambda(\lambda-1).$$

There are three distinct **eigenvalues** -1, 0, 1. Each has **algebraic multiplicity** 1.

Example 7.1.16. Find the real **eigenvalues** with their **algebraic multiplicities** for the matrix $A = \begin{pmatrix} 5 & -1 & 5 \\ 1 & -1 & -1 \\ -2 & 2 & 0 \end{pmatrix}$.

The **characteristic polynomial** of A is $det\begin{pmatrix} 5-\lambda & -1 & 5 \\ 1 & -1-\lambda & -1 \\ -2 & 2 & -\lambda \end{pmatrix} =$

$$(5-\lambda)(-1-\lambda)(-\lambda) + (-1)(-1)(-2) + (5)(1)(2) -$$
$$[(5-\lambda)(2)(-1) + (1)(-1)(-\lambda) + (-2)((-1-\lambda)(5)] =$$
$$-\lambda^3 + 4\lambda^2 + 5\lambda - 2 + 10 + 10 - 2\lambda - \lambda + 10 - 10\lambda =$$
$$-\lambda^3 + 4\lambda^2 - 8\lambda + 8 =$$
$$-(\lambda^3 - 4\lambda^2 + 8\lambda - 8) = -(\lambda - 2)(\lambda^2 - 2\lambda + 4).$$

The quadratic factor $\lambda^2 - 2\lambda + 4 = (\lambda - 1)^2 + 3$ has the complex roots $1 \pm i\sqrt{3}$ and no real roots. Therefore there is one real **eigenvalue**, 2, with **algbebraic multiplicity** 1.

Method 7.1.3. For each **eigenvalue** α of a square matrix A find a **basis** for the **eigenspace** E_α and the **geometric multiplicity** of the α.

For each real **eigenvalue** α of A find a **basis** for the **null space** of $A - \alpha I_n$. This will be a **basis** for the **$\alpha-$eigenspace**, E_α. The **dimension** of this space is the **geometric mutliplicity** of α.

Example 7.1.17. For each real **eigenvalue** α of the matrix $A = \begin{pmatrix} 6 & 4 \\ -4 & -2 \end{pmatrix}$ find a **basis** for its **eigenspace**. Then determine the **geometric multiplicity** of each real **eigenvalue**.

In **Example** (7.1.10) we found that this matrix has a unique real **eigenvalue**, 2, with algebraic multiplicity 2. We form the matrix $A - 2I_2$ and find a **basis** for its **null space**.

$A - 2I_2 = \begin{pmatrix} 4 & 4 \\ -4 & -4 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. In the usual way we get that $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is a **basis** for the **null space** of this matrix. Thus, $E_2 = Span\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)$. The **geometric multiplicity** of 2 is 1.

This example illustrates **Remark** (7.3) that the **geometric multiplicity** of an **eigenvalue** can be less than the **algebraic multiplicity**.

Example 7.1.18. Find **bases** for the real **eigenspaces** of the matrix $A = \begin{pmatrix} 3 & 5 \\ -1 & -3 \end{pmatrix}$. Determine the **geometric multiplicity** of each real **eigenvalue**.

In **Example** (7.1.11) we found that his matrix has real **eigenvalues** -2 and 2. For each of these **eigenvalues** we find a **basis** for the **eigenspaces**.

Computing the **eigenspace** for the **eigenvalue** -2:

$A-(-2)I_2 = A+2I_2 = \begin{pmatrix} 5 & 5 \\ -1 & -1 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. Then $E_{-2} = Span\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)$. The eigenvalue -2 has **geometric multiplicity** 1.

Computing the **eigenspace** for the **eigenvalue** 2:

$A-2I_2 = \begin{pmatrix} 1 & 5 \\ -1 & -5 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 5 \\ 0 & 0 \end{pmatrix}$. Then $E_2 = Span\left(\begin{pmatrix} -5 \\ 1 \end{pmatrix}\right)$. The **eigenvalue** 2 has **geometric multiplicity** 1.

Example 7.1.19. Find **bases** for the real **eigenspaces** of the matrix $A = \begin{pmatrix} 5 & -4 & 8 \\ -4 & 1 & -4 \\ -4 & 2 & -5 \end{pmatrix}$. Determine the **geometric multiplicity** of each real **eigenvalue**.

From **Example** (7.1.13) we know that this matrix has two distinct **eigenvalues**, -1 and 1.

Computing the **eigenspace** for the **eigenvalue** -1:

$A-(-1)I_3 = A+I_3 = \begin{pmatrix} 6 & -4 & 8 \\ -4 & 2 & -4 \\ -4 & 2 & -4 \end{pmatrix}$ The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$. The corresponding **homogeneous linear system** is

$$\begin{array}{ccccc} x & & & = & 0 \\ & y & - & 2z & = & 0 \end{array}.$$

There are two **leading variables** and one **free variable**, z. We set $z = t$ and get that the general solution consists of all vectors of the form $\begin{pmatrix} 0 \\ 2t \\ t \end{pmatrix}$ and consequently the vector $\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ is a **basis** for E_{-1}. The **geometric multiplicity** is 1.

Computing the **eigenspace** for the **eigenvalue** 1:

$A - I_3 = \begin{pmatrix} 4 & -4 & 8 \\ -4 & 0 & -4 \\ -4 & 2 & -6 \end{pmatrix}$. This matrix has **reduced echelon form**

$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. Now $E_1 = Span\left(\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}\right)$. The **geometric multiplicity** is 1 as contrasted with the **algebraic multiplicity** 2. We will learn in the next section that this means the matrix is not **similar** to a **diagonal matrix**.

Example 7.1.20. Find **bases** for the real **eigenspaces** of the matrix $A = \begin{pmatrix} 0 & 1 & 0 \\ -2 & 3 & 0 \\ -2 & 1 & 2 \end{pmatrix}$.

Determine the **geometric multiplicity** of each real **eigenvalue**.

From **Example** (7.1.14) we know that this matrix also has two distinct **eigenvalues**, 1 and 2. We find **bases** for the corresponding **eigenspaces**.

Computing the **eigenspace** for the **eigenvalue**1:

$A - I_3 = \begin{pmatrix} -1 & 1 & 0 \\ -2 & 2 & 0 \\ -2 & 1 & 1 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$.

From this we obtain that $E_1 = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right)$. The **geometric multiplicity** is 1.

Computing the **eigenspace** for the **eigenvalue** 2:

$A - 2I_3 = \begin{pmatrix} -2 & 1 & 0 \\ -2 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.
Now there are two **free variables**. In the usual way we obtain a **basis** for the **null space**, with is **basis** for the 2-eigen space: $E_2 = Span\left(\begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right)$. Thus, the eigenvalue 2 has **geometric multiplicity** 2. We shall see in the section that we can conlude from this information that the matrix A is **similar** to a **diagonal matrix**.

Example 7.1.21. Find **bases** for the real **eigenspaces** of the matrix $A = \begin{pmatrix} 5 & 2 & -4 \\ 6 & 3 & -5 \\ 10 & 4 & -8 \end{pmatrix}$.
Determine the **geometric multiplicity** of each real **eigenvalue**.

We found in **Example** (7.1.15) that this matrix has three distinct **eigenvalues**: -1, 0, 1. We proceed to find **bases** for their **eigenspaces**.

Computing the **eigenspace** for the **eigenvalue** -1:

$A - (-1)I_3 = A + I_3 = \begin{pmatrix} 6 & 2 & -4 \\ 6 & 4 & -5 \\ 10 & 4 & -7 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}$. Then $E_{-1} = Span\left(\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}\right)$.

Computing the **eigenspace** for the **eigenvalue** 0:

$A - 0I_3 = A = \begin{pmatrix} 5 & 2 & -4 \\ 6 & 3 & -5 \\ 10 & 4 & -8 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -\frac{2}{3} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 \end{pmatrix}$.

Then $E_0 = Span\left(\begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ 1 \end{pmatrix}\right)$.

Computing the **eigenspace** for the **eigenvalue** 1:

$A - I_3 = \begin{pmatrix} 4 & 2 & -4 \\ 6 & 2 & -5 \\ 10 & 4 & -9 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$.

Then $E_1 = Span\left(\begin{pmatrix} \frac{1}{2} \\ 1 \\ 1 \end{pmatrix}\right)$.

In each case the **geometric multiplicity** is one. We will see in the next section that if the **characteristic polynomial** of an $n \times n$ matrix A has n distinct real roots then the **geometric multiplicity** of each **eigenvalue** is also one. In this case the matrix A will be **similar** to a **diagonal matrix**.

Example 7.1.22. Find **bases** for the real **eigenspaces** of the matrix $A = \begin{pmatrix} 5 & -1 & 5 \\ 1 & -1 & -1 \\ -2 & 2 & 0 \end{pmatrix}$. Determine the **geometric multiplicity** of each real **eigenvalue**.

We found in **Example** (7.1.16) that this matrix has single real **eigenvalue**, 2, with **algebraic multiplicity** 1. We proceed to find a **basis** for the **eigenspace** E_2 of A :

$A - 2I_3 = \begin{pmatrix} 3 & -1 & 5 \\ 1 & -3 & -1 \\ -2 & 2 & -2 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$.

Therefore $E_2 = Span\left(\begin{pmatrix} -2 \\ -1 \\ 1 \end{pmatrix}\right)$ and the **eigenvalue** 2 has **geometric multiplicity** 1.

Exercises

In 1 - 4 verify that the each of the given vectors is an **eigenvector** for the given matrix. Determine the associated **eigenvalue** for each vector.

1. $A = \begin{pmatrix} -6 & 4 \\ -2 & 3 \end{pmatrix}, \boldsymbol{v}_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

2. $A = \begin{pmatrix} -7 & 4 \\ -4 & 1 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$

3. $A = \begin{pmatrix} -1 & 1 & -1 \\ -4 & -1 & 4 \\ -4 & 1 & 2 \end{pmatrix}, \boldsymbol{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$

4. $A = \begin{pmatrix} 3 & 7 & 7 \\ 1 & 3 & 1 \\ -1 & -7 & -5 \end{pmatrix}, \boldsymbol{v}_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$

In exercises 5 - 13 find the **characteristic polynomial** and **eigenvalues** of the given matrix. For each **eigenvalue** determine the **algebraic multiplicity**. See **Method** (7.1.1) and **Method** (7.1.2).

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

6. $A = \begin{pmatrix} -3 & 7 \\ -1 & 5 \end{pmatrix}.$

7. $A = \begin{pmatrix} 5 & 4 \\ -2 & -1 \end{pmatrix}.$

8. $A = \begin{pmatrix} 10 & -10 & 11 \\ 4 & -4 & 5 \\ -3 & 3 & -3 \end{pmatrix}$

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

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

11. $A = \begin{pmatrix} 1 & 3 & 6 \\ 0 & 4 & 6 \\ 0 & -3 & -5 \end{pmatrix}.$

12. $A = \begin{pmatrix} -6 & 11 & 6 \\ -2 & 4 & 2 \\ -4 & 7 & 3 \end{pmatrix}$.

13. $A = \begin{pmatrix} 3 & 5 & -5 \\ 2 & 3 & -2 \\ 2 & 5 & -4 \end{pmatrix}$.

In exercises 14 - 22 for each **eigenvalue** of the given matrix, find a **basis** for the associated **eigenspace** and determine the **geometric multiplicity** of the **eigenvalue**. See **Method** (7.1.3).

14. The matrix of exercise 5.

15. The matrix of exercise 6.

16. The matrix of exercise 7.

17. The matrix of exercise 8.

18. The matrix of exercise 9.

19. The matrix of exercise 10.

20. The matrix of exercise 11.

21. The matrix of exercise 12.

22. The matrix of exercise 13.

In exercises 23 - 28 answer true or false and give an explanation.

23. If $\boldsymbol{v}$ is an **eigenvector** for the matrix A with **eigenvalue** 2 and $\boldsymbol{w}$ is an eigenvector of A with eigenvalue 3 then $\boldsymbol{v} + \boldsymbol{w}$ is an eigenvector of A with eigenvalue 5.

24. If $\boldsymbol{v}, \boldsymbol{w}$ are distinct (not equal) **eigenvectors** of A with **eigenvalue** 2 then $\boldsymbol{v} - \boldsymbol{w}$ is an eigenvector of A with eigenvalue 0..

25. If $\boldsymbol{v}, \boldsymbol{w}$ are both **eigenvectors** of A with **eigenvalue** 2 then $\boldsymbol{v} - \boldsymbol{w}$ is an eigenvector of A with eigenvalue 2.

26. If $\boldsymbol{v}$ is an **eigenvector** of the matrix A with **eigenvalue** 3 then $4\boldsymbol{v}$ is an eigenvector of A with eigenvalue 12.

27. For a **non-invertible** $n \times n$ matrix A the **null space** of A is the 0 **eigenspace** of A.

28. If α is an **eigenvalue** of a matrix A with **geometric multiplicity** three then the **algebraic multiplicity** of A must be less than or equal three.

Challenge Exercises (Problems)

1. Let A be a square matrix. Prove that $\chi_A(\lambda) = \chi_{A^{Tr}}(\lambda)$.

2. Assume that $\boldsymbol{v}$ is an **eigenvector** for A with **eigenvalue** λ. Prove that $\boldsymbol{v}$ is an **eigenvector** of $A + cI_n$ with **eigenvalue** $\lambda + c$.

3. Let J_n be the $n \times n$ matrix with all 1's. Prove that J_n has **eigenvalue** n with **mulitiplicity** 1 and **eigenvalue** 0 with **geometric multiplicity** $n-1$.

4. Prove that $aJ_n + bI_n$ has **eigenvalue** $an + b$ with **geometric multiplicity** 1 and **eigenvalue** b with **geometric multiplicity** $n-1$.

5. Assume that $\boldsymbol{v}$ is an **eigenvector** of the matrix A with **eigenvalue** λ. Prove that $\boldsymbol{v}$ is a **eigenvector** of A^k with **eigenvalue** λ^k.

6. Assume that A and B are **similar matrices**, that is, there is an **invertible matrix** Q such that $B = QAQ^{-1}$. Prove that A and B have the same **characteristic polynomial**.

7. Suppose Q is **invertible**, $B = QAQ^{-1}$ and $\boldsymbol{v}$ is an **eigenvector** of A with **eigenvalue** λ. Prove that $Q\boldsymbol{v}$ is an **eigenvector** of B with **eigenvalue** λ.

8. Assume that the $n \times n$ matrices A and B commute, that is, $AB = BA$. Suppose $\boldsymbol{v}$ is an **eigenvector** for A with **eigenvalue** λ. Prove that $B\boldsymbol{v}$ is an **eigenvector** for A with **eigenvalue** λ.

9. Assume that $\boldsymbol{v}$ is an **eigenvector** of the **invertible matrix** A with **eigenvalue** λ. Prove that $\boldsymbol{v}$ is an **eigenvector** of A^{-1} with **eigenvalue** $\lambda^{-1} = \frac{1}{\lambda}$.

Quiz Solutions

1. $det \begin{pmatrix} 2 & 4 & 7 \\ 2 & 1 & -3 \\ 5 & 2 & 6 \end{pmatrix} =$

$$2 \times 1 \times 6 + 2 \times 2 \times 7 + 5 \times 4 \times (-3) - [5 \times 1 \times 7 + 2 \times 4 \times 6 + 2 \times 2 \times (-3)] =$$

$$(12 + 28 - 60) - (35 + 48 - 12) = -20 - 71 =$$

$$-91$$

Not right, see **determinant of a 3 x 3 matrix**.

2. $det \begin{pmatrix} 2-\lambda & 7 \\ -2 & -1-\lambda \end{pmatrix} = (-2-\lambda)(-1-\lambda) + 14 = \lambda^2 - \lambda + 12$

Not right, see **determinant of a 2 x 2 matrix**.

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

$$(3-\lambda)(3-\lambda)(-\lambda) - 2 - 4 + 2(3-\lambda) + 2(3-\lambda) + 2\lambda =$$
$$-\lambda^3 + 6\lambda^2 - 11\lambda + 6$$

Not right, see **determinant of a 3 x 3 matrix**.

4. a) $\lambda^2 + 3\lambda - 10 = (\lambda+5)(\lambda-2)$

Not right, see a high school intermediate algebra text.

b) $\lambda^3 - 7\lambda + 6 = (\lambda-1)(\lambda-2)(\lambda+3)$

Not right, see a high school intermediate algebra text

5. The **reduced echelon form** of the matrix A is $R = \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

The **homogeneous linear system** with **coefficient matrix** R is

$$\begin{array}{ccccccccc} x_1 & & & - & x_3 & + & 2x_4 & = & 0 \\ & & x_2 & + & x_3 & - & x_4 & = & 0 \end{array}$$

There are two **free variables**, x_3, x_4. We set these equal to parameters and solve for all the variables in terms of these parameters. We get

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

$\left(\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix}\right)$ is a **basis** for the **null space** of A.

Not right, see **Method** (3.2.2).

7.2. Diagonalization of Matrices

We find conditions which guarantee that a **square matrix** is **similar** to a **diagonal matrix**. These will depend on the **characteristic polynomial** of the matrix and the its **eigenvalues**.

Am I Ready for This Material

The following are those concepts which are used extensively in this section:

null space of a matrix

diagonal matrix

linearly independent sequence of vectors

basis of $\mathbb{R}^n$

change of basis matrix

determinant of a square matrix

similar matrices

eigenvector of a square matrix

eigenvalue of a square matrix

characteristic polynomial of a square matrix

characteristic equation of a square matrix

algebraic multiplicity of an eigenvalue of a square matrix

eigenspace of a square matrix

geometric multiplicity of an eigenvalue of a square matrix

Quiz

For each of the following matrices

a) Find the **characteristic equation**;

b) Solve the **characteristic equation** to get the **eigenvalues** and determine their **algebraic multiplicity**;

c) Find a **basis** for each **eigenspace**; and

d) Determine the **geometric multiplicity** of each **eigenvalue**.

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

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

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

Quiz Solutions

New Concepts

Two especially important concepts are introduced here for square matrices

diagonalizable matrix

diagonalizing matrix

Theory (Why It Works)

Let $T : V \to V$ is a **linear transformation**, where V is a **finite dimensional vector space**. Assume that $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n), \mathcal{B}' = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$ are **bases** for V, A the **matrix** of T with respect to $\mathcal{B}$ and A' the matrix of T with respect to $\mathcal{B}'$. We have previously seen that there is an **invertible matrix** P such that $A' = P^{-1}AP$.

In fact, the matrix P is the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$, that is, the matrix whose columns consists of the **coordinate vectors** of the vectors $\boldsymbol{w}_j$ in terms of the $\boldsymbol{v}_i$. We previously gave a special name to this relation between square matrices:

Definition 6.13

Let A, B be $n \times n$ matrices. We say that B is ***similar*** to A if there is an **invertible matrix** P such that $P^{-1}AP = B$ or, equivalently, $AP = PB$.

We further recall the definition of a **diagonal matrix**: it is a square matrix such that for all $i \neq j$ the $(i,j)-$entry is zero, that is, the entries off the **main diagonal** are all zero.

Example 7.2.1.

$$\begin{pmatrix} 2 & 0 \\ 0 & -5 \end{pmatrix}, \begin{pmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -7 \end{pmatrix}.$$

Diagonal matrices act, via multiplication, in a very simple way: If

$$D = diag(\alpha_1, \alpha_2, \dots, \alpha_n)$$

and A is a matrix with rows $\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_n$, $A = \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_n \end{pmatrix}$, then $DA = \begin{pmatrix} \alpha_1\boldsymbol{r}_1 \\ \alpha_2\boldsymbol{r}_2 \\ \vdots \\ \alpha_n\boldsymbol{r}_n \end{pmatrix}$, that is, the i^{th} row of DA is α_i times the i^{th} row of A.

On the other hand, if $A = \begin{pmatrix} \boldsymbol{c}_1 & c_2 & \dots & \boldsymbol{c}_n \end{pmatrix}$, where the $\boldsymbol{c}_j$ are the columns of A ,then $AD = \begin{pmatrix} \alpha_1\boldsymbol{c}_1 & \alpha_2\boldsymbol{c}_2 & \dots & \alpha_n\boldsymbol{c}_n \end{pmatrix}$, so the j^{th} column of AD is α_j times the j^{th} column of A.

Using **diagonal matrices** can simplify computations since powers of a diagonal matrix, the **determinant**, and the **inverse** (when invertible) are all exceedingly easy to compute.

Example 7.2.2. Let $A = diag(2, -3, \frac{2}{3}) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & \frac{2}{3} \end{pmatrix}$.

Then $det(A) = 2 \times (-3) \times \frac{2}{3} = -4$.

On the other hand, $A^3 = diag(8, -27, \frac{8}{27})$, $A^4 = diag(16, 81, \frac{16}{81})$ and $A^{-1} = diag(\frac{1}{2}, -\frac{1}{3}, \frac{3}{2})$.

Suppose that a square matrix A is not itself **diagonal** but is **similar** to a diagonal matrix, say $P^{-1}AP = D = diag(d_1, d_2, \dots, d_n)$. Then setting $Q = P^{-1}$ we get that $A = Q^{-1}DQ$. Note that $(Q^{-1}DQ)^m = Q^{-1}D^mQ$.

Also, A is **invertible** if and only if D is invertible, if and only if $d_i \neq 0, 1 \leq i \leq n$. In this case $A^{-1} = (Q^{-1}DQ)^{-1} = Q^{-1}D^{-1}Q = Q^{-1}diag(\frac{1}{d_1}, \frac{1}{d_2}, \ldots, \frac{1}{d_n})Q$. In this way, all kinds of computations involving A can be simplified. This motivates the following definition:

Definition 7.6. We will say that a square matrix A is ***diagonalizable*** if A is **similar** to a **diagonal matrix**. This means that there is an **invertible matrix** P and diagonal matrix $D = diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$ such that $P^{-1}AP = D$ or, what amounts to the same thing, $AP = PD$. Any diagonal matrix D which is **similar** to A will be called a ***diagonal form*** of A. Finally, a matrix P such that $P^{-1}AP$ is diagonal is called a ***diagonalizing matrix***.

Example 7.2.3. Let $A = \begin{pmatrix} 5 & -2 \\ 3 & 0 \end{pmatrix}, P = \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix}$.

Then $AP = \begin{pmatrix} 4 & 3 \\ 6 & 3 \end{pmatrix} = P\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$.

Example 7.2.4. Let $A = \begin{pmatrix} -2 & -2 & -2 \\ -2 & 1 & 4 \\ 3 & -6 & -9 \end{pmatrix}, P = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 0 & -2 \\ 0 & 1 & 3 \end{pmatrix}$.

Then $AP = \begin{pmatrix} -6 & -6 & -4 \\ -3 & 0 & 8 \\ 0 & -3 & -12 \end{pmatrix} = P\begin{pmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -4 \end{pmatrix}$.

Remark 7.7. (1) In the definition of a **diagonalizable matrix** neither the diagonalizing matrix P nor the diagonal form D is unique. First of all, we can take scalar multiples of the columns of P. This will not change D. However, we can also permute the columns of P and this will permute the columns of D.

(2) Not every matrix is diagonalizable. To see this consider $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Suppose that $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, D = \begin{pmatrix} e & 0 \\ 0 & f \end{pmatrix}$ are such that $AP = PD$. Then $\begin{pmatrix} c & d \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} ae & bf \\ ce & df \end{pmatrix}$. Therefore $ce = 0 = df$. Now it is not the case that both c and d are equal to zero since P is **invertible**. Suppose $c \neq 0$. Then $e = 0$. But $c = ae$ and therefore $c = 0$. so we must have $c = 0$. In a similar fashion we get a contradiction if $d \neq 0$. Therefore, also $d = 0$ but as stated immediately above this is not possible. Therefore, A is not diagonalizable.

Assume that A is diagonalizable to $D = diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$ via P so that $P^{-1}AP = D$. Let $\boldsymbol{c}_j$ be the j^{th} column of P. Then we have

$$AP = A(\boldsymbol{c}_1\ \boldsymbol{c}_2\ \ldots\ \boldsymbol{c}_n) =$$

$$(A\boldsymbol{c}_1\ A\boldsymbol{c}_2\ \ldots\ A\boldsymbol{c}_n) = PD =$$

$$(\boldsymbol{c}_1\ \boldsymbol{c}_2\ \ldots\ \boldsymbol{c}_n)diag(\alpha_1\ \alpha_2\ \ldots\ \alpha_n) =$$

$$(\alpha_1\boldsymbol{c}_1\ \alpha_2\boldsymbol{c}_2\ \ldots\ \alpha_n\boldsymbol{c}_n).$$

Therefore the columns of P are **eigenvectors** of A and the **eigenvalue** corresponding to $\boldsymbol{c}_j$ is α_j.

The columns of P form a **basis** of $\mathbb{R}^n$ since P is **invertible**. On the other hand, suppose A has n **eigenvectors** $\boldsymbol{v}_j, 1 \leq j \leq n$, with corresponding eigenvalue α_j such that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is **linearly independent**. Then we claim that A is diagonalizable via $P = \begin{pmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n\end{pmatrix}$ to $diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

That P is **invertible** follows from the fact that P is an $n \times n$ matrix and the sequence of its columns is **linearly independent**. Also,

$$AP = A(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_n) =$$

$$(A\boldsymbol{v}_1\ A\boldsymbol{v}_2\ \ldots\ A\boldsymbol{v}_n) =$$

$$(\alpha_1\boldsymbol{v}_1\ \alpha_2\boldsymbol{v}_2\ \ldots\ \alpha_n\boldsymbol{v}_n)$$

since $\boldsymbol{v}_j$ is an **eigenvector with eigenvalue α_j**. However, $(\alpha_1\boldsymbol{v}_1\ \alpha_2\boldsymbol{v}_2\ \ldots\ \alpha_n\boldsymbol{v}_n) = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_n)diag(\alpha_1\ \alpha_2\ \ldots\ \alpha_n) = PD$. Thus, $AP = PD$ which implies that $P^{-1}AP = D$. We have therefore proved:

Theorem 7.2.1. *An $n \times n$ matrix A is* **diagonalizable** *if and only if some sequence of* **eigenvectors** *of A is a* **basis** *of $\mathbb{R}^n$.*

Before expliciting giving a recipe for determining whether a square matrix A is **diagonalizable** and, if it is, how to actually find a **diagonal form** and a **diagonalizing matrix** we first discuss some criteria for diagonalization.

Theorem 7.2.2. *Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ be* **eigenvectors** *of the $n \times n$ matrix A with distinct eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_m$. Then $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$ is* **linearly independent**.

Proof. Before giving the general argument we rehearse it for the case that $m = 3$ and the eigenvalues are 2,3,4. Assume that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \boldsymbol{0}_n$ is a **dependence relation**. We must show that $c_1 = c_2 = c_3 = 0$. We first prove that $c_3 = 0$. Consider the matrix $B = (A - 2I_n)(A - 3I_n)$. Note that by **Remark** (7.1) $\boldsymbol{v}_1$ is a **eigenvector** of $A - 3I_n$ with **eigenvalue** -1 and in the **null space** of $A - 2I_n$. Therefore $B\boldsymbol{v}_1 = \boldsymbol{0}_n$. In a similar fashion $B\boldsymbol{v}_2 = \boldsymbol{0}_n$. On the other hand, $\boldsymbol{v}_3$ is an eigenvector of $A - 2I_n$ with eigenvalue 2 and an eigenvector of $A - 3I_n$ with eigenvalue 1. Therefore $B\boldsymbol{v}_3 = (2)(1)\boldsymbol{v}_3 = 2\boldsymbol{v}_3$. However, since $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \boldsymbol{0}_n$ it follows that

$$B(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3) = 2c_3\boldsymbol{v}_3 = \boldsymbol{0}_n$$

and, consequently, $c_3 = 0$ since $\boldsymbol{v}_3 \neq \boldsymbol{0}_n$. Thus, the dependence relation is $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 = \boldsymbol{0}_n$. Now set $C = A - 2I_n$. Then $C\boldsymbol{v}_1 = \boldsymbol{0}_n$ and $C\boldsymbol{v}_2 = \boldsymbol{v}_2$. Multiplying the dependence relation $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 = \boldsymbol{0}_n$ by C we get

$$C(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_2\boldsymbol{v}_2 = \boldsymbol{0}_n$$

and therefore $c_2 = 0$ since $\boldsymbol{v}_2 \neq \boldsymbol{0}_n$. Now the dependence relation is $c_1\boldsymbol{v}_1 = \boldsymbol{0}_n$. Since $\boldsymbol{v}_1 \neq \boldsymbol{0}_n$ we can conclude that $c_1 = 0$.

We now go on to the general case. Assume to the contrary that $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_m)$ is **linearly dependent** so that there is a **non-trivial dependence relation** $c_1\boldsymbol{v}_1 + \cdots + c_m\boldsymbol{v}_m = \boldsymbol{0}_n$. Let l be the maximum index such that $c_l \neq 0$. We will get a contradiction. Set $A_i = A - \alpha_i I_n$. Then by **Remark** (7.1) for $j \neq i, A_i\boldsymbol{v}_j = (\alpha_j - \alpha_i)\boldsymbol{v}_j$ and $A_i\boldsymbol{v}_i = \boldsymbol{0}_n$. Now set $D = A_1A_2\dots A_{l-1}$. Then $D\boldsymbol{v}_j = \boldsymbol{0}_n$ for $j < l$ and $D\boldsymbol{v}_l = (\alpha_l - \alpha_1)(\alpha_l - \alpha_2)\dots(\alpha_l - \alpha_{l-1})\boldsymbol{v}_l \neq \boldsymbol{0}_n$. Applying D to $c_1\boldsymbol{v}_1 + \dots c_l\boldsymbol{v}_l$ we get

$$D(c_1\boldsymbol{v}_1 + \cdots + c_l\boldsymbol{v}_l) = c_1D\boldsymbol{v}_1 + \cdots + c_lD\boldsymbol{v}_l =$$

$$\boldsymbol{0}_n + \cdots + c_l(\alpha_l - \alpha_1)\dots(\alpha_l - \alpha_{l-1})\boldsymbol{v}_l =$$

$$c_l(\alpha_l - \alpha_1)\dots(\alpha_l - \alpha_{l-1})\boldsymbol{v}_l.$$

However, since $c_1\boldsymbol{v}_1 + \cdots + c_l\boldsymbol{v}_l = \boldsymbol{0}_n$ we have $D(c_1\boldsymbol{v}_1 + \cdots + c_l\boldsymbol{v}_l) = \boldsymbol{0}_n$. Therefore

$$c_l(\alpha_l - \alpha_1)\dots(\alpha_l - \alpha_{l-1})\boldsymbol{v}_l = \boldsymbol{0}_n.$$

Since $\boldsymbol{v}_l \neq \boldsymbol{0}_n$ it follows that $c_l(\alpha_l - \alpha_1)\dots(\alpha_l - \alpha_{l-1}) = 0$. Since $(\alpha_l - \alpha_1)\dots(\alpha_l - \alpha_{l-1}) \neq 0$ we can conclude that $c_l = 0$, which contradicts our assumption. It therefore follows that $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_m)$ is **linearly independent**. $\square$

We make use of Theorem (7.2.2) in taking a further step towards developing useful criteria for determining if a square matrix is **diagonalizable**.

Theorem 7.2.3. *Let $\alpha_1, \alpha_2, \dots, \alpha_k$ be distinct* **eigenvalues** *of the $n \times n$ matrix A. Let $\mathcal{B}_i$ be a basis for E_{α_i}, the α_i-eigenspace. Set $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_k$. Then $\mathcal{B}$ is* **linearly independent**.

Proof. We sketch the proof in the case $k = 3$ and for the purposes of exposition we assume that $dim(E_{\alpha_1}) = 2, dim(E_{\alpha_2}) = 3$ and $dim(E_{\alpha_3}) = 3$.

Let $\mathcal{B}_1 = (\boldsymbol{u}_1, \boldsymbol{u}_2), \mathcal{B}_2 = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$, and $\mathcal{B}_3 = (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$. Suppose

$$c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + d_1\boldsymbol{v}_1 + d_2\boldsymbol{v}_2 + d_3\boldsymbol{v}_3 + e_1\boldsymbol{w}_1 + e_2\boldsymbol{w}_2 + e_3\boldsymbol{w}_3 = \boldsymbol{0}_n \tag{7.2}$$

is a **dependence relation**. We need to show all the coefficients are zero:

$$c_1 = c_2 = d_1 = d_2 = d_3 = e_1 = e_2 = e_3 = 0.$$

Set $\boldsymbol{u} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2$. Since $\boldsymbol{u}_1, \boldsymbol{u}_2 \in E_{\alpha_1}$, a **subspace** of $\mathbb{R}^n$, it follows that $\boldsymbol{u} \in E_{\alpha_1}$ and therefore $A\boldsymbol{u} = \alpha_1\boldsymbol{u}$. In a similar fashion, if we set $\boldsymbol{v} = d_1\boldsymbol{v}_1 + d_2\boldsymbol{v}_2 + d_3\boldsymbol{v}_3$, and $\boldsymbol{w} = e_1\boldsymbol{w}_1 + e_2\boldsymbol{w}_2 + e_3\boldsymbol{w}_3$ then $A\boldsymbol{v} = \alpha_2\boldsymbol{v}$ and $A\boldsymbol{w} = \alpha_3\boldsymbol{w}$.

If $\boldsymbol{u} \neq \boldsymbol{0}_n$ then $\boldsymbol{u}$ is an **eigenvector with eigenvalue** α_1. In a similar fashion if $\boldsymbol{v} \neq \boldsymbol{0}_n$ then $\boldsymbol{v}$ is an **eigenvector with eigenvalue** α_2 and if $\boldsymbol{w} \neq \boldsymbol{0}_n$ then $\boldsymbol{w}$ is an eigenvector with eigenvalue α_3.

It follows from (7.2) and the definition of $\boldsymbol{u}, \boldsymbol{v}$ and $\boldsymbol{w}$ that $\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w} = \boldsymbol{0}_n$, a **non-trivial dependence relation**. It now is a consequence of **Theorem** (7.2.2) that $\boldsymbol{u} = \boldsymbol{v} = \boldsymbol{w} = \boldsymbol{0}_n$.

Thus, $c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 = \boldsymbol{0}_n$. However, since $(\boldsymbol{u}_1, \boldsymbol{u}_2)$ is a basis for E_{α_1} this sequence is **linearly independent**. Therefore $c_1 = c_2 = 0$. In exactly the same fashion we conclude from $d_1\boldsymbol{v}_1 + d_2\boldsymbol{v}_2 + d_3\boldsymbol{v}_3 = \boldsymbol{0}_n$ that $d_1 = d_2 = d_3 = 0$ and that also $e_1 = e_2 = e_3 = 0$. Thus, all the coefficients are zero and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent**. $\square$

A consequence of the last result is the following criteria for a square matrix A to be **diagaonalizable**.

Theorem 7.2.4. *Suppose that $\alpha_1, \alpha_2, \dots, \alpha_k$ are all the distinct* **eigenvalues** *of an $n \times n$ matrix A. Let $\mathcal{B}_i$ be a* **basis** *for the* **eigenspace** *E_{α_i} for $i = 1, 2, \dots, k$ and set $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_k$. Then A is* **diagonalizable** *if and only if the sequence $\mathcal{B}$ has length n.*

If for some i the **geometric multiplicity** of α_i is less then the **algebraic multiplicity** then it will not be possible to find a **basis** consisting of **eigenvectors**. Therefore we have following more explicit test for **diagonalizability**:

Theorem 7.2.5. *The $n \times n$ matrix A is* **diagonalizable** *if and only if the* **characteristic polynomial** *factors into linear factors and the* **geometric multiplicity** *of each* **eigenvalue** *is equal to its* **algebraic multiplicity**.

Of course, if the **algebraic multiplicity** of an **eigenvalue** is one then the **geometric multiplicity** of that eigenvalue must also be one. This observation implies the following:

Corollary 7.2.6. *Let A be an $n \times n$ matrix. If the* **characteristic polynomial** *has n distinct (real) linear factors then the matrix A is* **diagonalizable**. *Equivalently, if A has n distinct real* **eigenvalues** *then A is* **diagonalizable**.

Implicit in **Theorem** (7.2.5) is a method for determining whether a matrix is **diagonalizable**. We break it down into steps:

DIAGONALIZING A SQUARE MATRIX A

Assume A is an $n \times n$ matrix.

1) Use **Method** (7.1.1) to find the **characteristic polynomial** $\chi_A(\lambda)$ of A. If $\chi_A(\lambda)$ has an irreducible quadratic factor, **STOP**. This matrix is not **diagonalizable** by a real matrix. Otherwise,

2) Apply **Method** (7.1.2) to determine the **eigenvalues** of A and their **algebraic multiplicity**. (Note that the sum of the **algebraic multiplicities** will be equal to n.)

3) For each **eigenvalue** α find a **basis** $\mathcal{B}_\alpha$ for the **eigenspace** E_α using **Method** (7.1.3) and use this to compute the **geometric multiplicity** of α.

If the **geometric multiplicity** of some **eigenvalue** α is less than its **algebraic multiplicity** then **STOP**. This matrix is not **diagonalizable**. Otherwise continue.

4) Assume for each α the **algebraic multiplicity** of α is equal to the **geometric multiplicity** of α. Then the matrix is **diagonalizable** to a diagonal matrix D whose diagonal entries are the **eigenvalues** of A, each occuring as many times as its **algebraic multiplicity**. A **diagonalizing matrix** P is obtained as follows:

5) Let $\mathcal{B}_\alpha$ be the **basis** of the α **eigenspace**, E_α, for each **eigenvalue** α obtained in 3) and set $\mathcal{B} = \mathcal{B}_{\alpha_1} \cup \mathcal{B}_{\alpha_2} \cup \cdots \cup \mathcal{B}_{\alpha_k}$ where the α_i are taken in the same order as they occur in the matrix D. Now let P be the matrix whose sequence of columns is equal to $\mathcal{B}$. Then P is a **diagonalizing matrix**.

Example 7.2.5. Let $A = \begin{pmatrix} -3 & 6 \\ 1 & 2 \end{pmatrix}$.

The **characteristic polynomial** is $\chi_A(\lambda) = det \begin{pmatrix} -3-\lambda & 6 \\ 1 & 2-\lambda \end{pmatrix} =$

$$(-3-\lambda)(2-\lambda) - 6 = \lambda^2 + \lambda - 12 = (\lambda+4)(\lambda-3).$$

Setting $\chi_A(\lambda) = 0$ we get the **eigenvalues** are $\lambda = -4, 3$. Since these are distinct (each has **algebraic multiplicity** one) the matrix is **diagonalizable**. We proceed to find **eigenvectors** and an **invertible matrix** P such that $P^{-1}AP$ is diagonal.

Computing eigenvectors for the eigenvalue $\lambda = -4$:

$A - (-4)I_2 = \begin{pmatrix} 1 & 6 \\ 1 & 6 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 6 \\ 0 & 0 \end{pmatrix}$.

This gives rise to the equation $x_1 + 6x_2 = 0$.

Setting the **free variable** x_2 equal to t we get $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -6t \\ t \end{pmatrix} = t\begin{pmatrix} -6 \\ 1 \end{pmatrix}$. The vector $\begin{pmatrix} -6 \\ 1 \end{pmatrix}$ **spans** E_{-4}.

Computing eigenvectors for the eigenvalue $\lambda = 3$:

$A - 3I_2 = \begin{pmatrix} -6 & 6 \\ 1 & -1 \end{pmatrix}$ The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$.

This gives the equation $x_1 - x_2 = 0$.

Setting the **free variable** x_2 equal to t we get $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} t \\ t \end{pmatrix} = t\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ **spans** E_3. Set $P = \begin{pmatrix} -6 & 1 \\ 1 & 1 \end{pmatrix}$.

Then $AP = \begin{pmatrix} -3 & 6 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -6 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 24 & 3 \\ -4 & 3 \end{pmatrix} = \begin{pmatrix} -6 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} -4 & 0 \\ 0 & 3 \end{pmatrix}$.

Example 7.2.6. Let $A = \begin{pmatrix} 3 & 8 \\ -2 & -5 \end{pmatrix}$.

The **characteristic polynomial** is $\chi_A(\lambda) = det(\begin{pmatrix} 3-\lambda & 8 \\ -2 & -5-\lambda \end{pmatrix}) =$

$$(3-\lambda)(-5-\lambda)+16 = \lambda^2+2\lambda+1 = (\lambda+1)^2.$$

Therefore, $\lambda = -1$ is an **eigenvalue** with **algebraic multiplicity** 2. We find **basis** of the (-1)-**eigenspace**, E_{-1}.

$A-(-1)I_2 = A+I_2 = \begin{pmatrix} 4 & 8 \\ -2 & -4 \end{pmatrix}$. This matrix has **reduced echelon form**

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

There are two variables, one **leading variable** and therefore one **free variable**. Therefore the **geometric multiplicity** of the eigenvalue -1 is 1 which is less than the **algebraic multiplicity**. Consequently, this matrix is not **diagonalizable**.

Example 7.2.7. Let $A = \begin{pmatrix} 8 & 3 & 6 \\ -4 & -1 & -4 \\ -5 & -3 & -3 \end{pmatrix}$.

The **characteristic polynomial** is $\chi_A(\lambda) = det\begin{pmatrix} 8-\lambda & 3 & 6 \\ -4 & -1-\lambda & -4 \\ -5 & -3 & -3-\lambda \end{pmatrix} =$

$$(8-\lambda)(-1-\lambda)(-3-\lambda)+60+72+30(-1-\lambda)+12(-3-\lambda)-12(8-\lambda) =$$

$$-\lambda^3+4\lambda^2+29\lambda+24+60+72-30-30\lambda-96+12\lambda-36-12\lambda =$$

$$-\lambda^3+4\lambda^2-\lambda-6 = -(\lambda+1)(\lambda-2)(\lambda-3).$$

Thus the **eigenvalues** are -1,2,3. They are distinct (all have **algebraic multiplicity** 1). Therefore the matrix is **diagonalizable**. We proceed to find a matrix P which will diagonalize A by finding an **eigenvector** for each **eigenvalue**.

Computing eigenvectors for the eigenvalue $\lambda = -1$:

We form the matrix $A-(-1)I_3 = A+I_3$ and find a **basis** for it **null space**.

$A+I_3 = \begin{pmatrix} 9 & 3 & 6 \\ -4 & 0 & -4 \\ -5 & -3 & -2 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$.

This is the **coefficient matrix** for the **homogeneous linear system**

$$\begin{array}{ccccccc} x_1 & & & + & x_3 & = & 0 \\ & & x_2 & - & x_3 & = & 0 \end{array}.$$

There are two **leading variables** (x_1, x_2) and one **free variable** (x_3). Setting the free variable $x_3 = t$ we get $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -t \\ t \\ t \end{pmatrix} = t\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ and therefore $E_{-1} = Span\left(\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}\right)$.

Computing eigenvectors for the eigenvalue $\lambda = 2$:

$A-2I_3 = \begin{pmatrix} 6 & 3 & 6 \\ -4 & -3 & -4 \\ -5 & -3 & -5 \end{pmatrix}$ The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

This is the **coefficient matrix** for the **homogeneous linear system**

$$\begin{array}{ccccccc} x_1 & & & + & x_3 & = & 0 \\ & & x_2 & & & = & 0 \end{array}.$$

There are two **leading variables** (x_1, x_2) and one **free variable** (x_3). Setting $x_3 = t$ and solving for x_1, x_2 we get $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -t \\ 0 \\ t \end{pmatrix} = t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$. Thus, $E_2 = Span\left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right)$.

Finally we find a **basis** of the **eigenspace** for $\lambda = 3$:

$A-3I_3 = \begin{pmatrix} 5 & 3 & 6 \\ -4 & -4 & -4 \\ -5 & -3 & -6 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & \frac{3}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}$.

This is the **coefficient matrix** for the **homogeneous linear system**

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

There are two **leading variables** (x_1, x_2) and one **free variable** (x_3). We set the free variable x_3 equal to a parameter t and get $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -\frac{3}{2}t \\ \frac{1}{2}t \\ t \end{pmatrix} = t \begin{pmatrix} -\frac{3}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}$.

In order to "clear fractions" we take $t = 2$ and get that the vector $\begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix}$ which **spans** E_3.

Now we can make a **diagonalizing matrix** $P = \begin{pmatrix} -1 & -1 & -3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ and $P^{-1}AP = diag\{-1, 2, 3\}$.

Remark 7.8. All the definitions and theorems of this section apply to matrices with entries in a **field** other than the reals, $\mathbb{R}$, in particular, to the complex numbers, $\mathbb{C}$. Since $\mathbb{C}$ is algebraically closed, the **characteristic polynomial** of an $n \times n$ complex matrix A will always have n roots (counting multiplicities). The matrix will be diagonalizable (via a complex matrix) if and only if there is a basis for $\mathbb{C}^n$ consisting of eigenvectors for A if and only if for each eigenvalue α the algebraic multiplicity of α is equal to the geometric multiplicity.

Diagonalizable Linear Operators

The definitions and theorems of this section can be extended to a **linear operator** T on a **finite dimensional vector space** V. For example,

Definition 7.7. We will say that a **linear operator** $T : V \to V$, where V is a **finite dimensional vector space**, is ***diagonalizable*** if there exists a **basis** $\mathcal{B}$ for V such that the **matrix** of T with respect to $\mathcal{B}$ is a **diagonal matrix**.

The following is almost an immediate consequence of the definition of a diagonalizable linear operator:

Theorem 7.2.7. *Let V be a* **finite dimensional vector space** *and let $T : V \to V$ be a* **linear operator**. *Then T is diagonalizable if and only if there is a* **basis** *$\mathcal{B}$ for V consisting of* **eigenvectors** *for T.*

Proof. Assume that T is diagonalizable. Let $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ be a **basis** for V such that the **matrix** of T with respect to $\mathcal{B}$ is the diagonal matrix $A = diag(d_1, d_2, \ldots, d_n)$. This means that

$$T(\boldsymbol{v}_j) = 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_{j-1} + d_j\boldsymbol{v}_j + 0\boldsymbol{v}_{j+1} + \cdots + 0\boldsymbol{v}_n \tag{7.3}$$

Thus, $\boldsymbol{v}_j$ is an **eigenvector** for T with **eigenvalue** d_j. Therefore the **basis** $\mathcal{B}$ consists of **eigenvectors** for T.

Conversely, assume that there is a **basis** $\mathcal{B} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ consisting of **eigenvectors** for T. Let $T(\boldsymbol{v}_j) = d_j\boldsymbol{v}_j$. This means that $[T(\boldsymbol{v}_j)]_{\mathcal{B}} = d_j\boldsymbol{e}_j$ and therefore the **matrix** of T with respect to the **basis** $\mathcal{B}$ is the diagonal matrix $diag(d_1, d_2, \ldots, d_n)$. □

Let $T : V \to V$ be a **linear operator** of a **finite dimensional vector space** V and let $\mathcal{B}$ be any **basis** for V and $A = \mathcal{M}_T(\mathcal{B}, \mathcal{B})$, the **matrix** of T with respect to $\mathcal{B}$. In order for T to be diagonalizable it is not necessary that A be a **diagonal matrix**, but rather A is a **diagonalizable matrix** as we show in the next theorem.

Theorem 7.2.8. *Let V be a* **finite dimensional vector space**, *$\mathcal{B}$ be a* **basis** *of V, and $T : V \to V$ be a* **linear operator** *on V. Let A be the* **matrix** *of T with respect to $\mathcal{B}$. Then T is diagonalizable if and only if A is* **diagonalizable**.

Proof. Assume T is **diagonalizable**. Let $\mathcal{B}'$ be a **basis** for V such that the **matrix** A' of T with respect to $\mathcal{B}'$ is diagonal.

Set $P = P_{\mathcal{B}' \to \mathcal{B}}$, the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$. Then by **Theorem** (6.3.2) $A' = P^{-1}AP$. This is precisely what it means for the matrix A to be **diagonalizable**.

Conversely, assume that A is **diagonalizable**. This means that there is an **invertible matrix** P such that $A' = P^{-1}AP$ is a diagonal matrix. Denote the j^{th} column of P by $\boldsymbol{p}_j$.

Let $\boldsymbol{v}'_j$ be the vector in V with $[\boldsymbol{v}'_j]_{\mathcal{B}} = \boldsymbol{p}_j$. Since the matrix P is **invertible**, in particular, the sequence of its columns is **linearly independent**. It then follows from **Theorem** (5.4.2) that the sequence $(\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n)$ is **linearly independent**. Since the dimension of V is n (there are n vectors in the **basis** $\mathcal{B}$) from **Theorem** (5.3.8) we can conclude that $\mathcal{B}' = (\boldsymbol{v}'_1, \boldsymbol{v}'_2, \ldots, \boldsymbol{v}'_n)$ is a **basis** for V. Moreover, P is the **change of basis matrix** from $\mathcal{B}'$ to $\mathcal{B}$. By **Theorem** (6.3.2) the **matrix** of T with respect to $\mathcal{B}'$ is $A' = P^{-1}AP$, a **diagonal matrix**. Thus, T is a **diagonalizable linear operator**. □

What You Can Now Do

1. For a square matrix A determine if A is **diagonalizable**.

2. If the $n \times n$ matrix A is **diagonalizable** find a **basis** for $\mathbb{R}^n$ consisting of **eigenvectors** for A.

3. If the $n \times n$ matrix A is **diagonalizable**, find a **diagonalizing matrix** and a diagonal form for A.

Method (How To Do It)

Method 7.2.1. All the things you are expected to do for this section are contained in the following procedure.

DIAGONALIZING A SQUARE MATRIX A

Assume A is an $n \times n$ matrix.

1) Use **Method** (7.1.1) to find the **characteristic polynomial** $\chi_A(\lambda)$ of A. If $\chi_A(\lambda)$ has an irreducible quadratic factor, **STOP**. This matrix is not **diagonalizable** by a real matrix. Otherwise,

2) Apply **Method** (7.1.2) to determine the **eigenvalues** of A and their **algebraic multiplicity**. (Note that the sum of the **algebraic multiplicities** will be equal to n.)

3) For each eigenvalue α find a **basis**, $\mathcal{B}_\alpha$, for the the **eigenspace** E_α using **Method** (7.1.3) and use this to determine the **geometric multiplicity** of α.

If the **geometric multiplicity** of some **eigenvalue** α is less than its **algebraic multiplicity** then **STOP**. This matrix is not **diagonalizable**. Otherwise continue.

4) Assume for each α the **algebraic multiplicity** of α is equal to the **geometric multiplicity** of α. Then the matrix is **diagonalizable** to a diagonal matrix D whose diagonal entries are the **eigenvalues** of A, each occuring as many times as its **algebraic multiplicity**. A **diagonalizing matrix** P is obtained as follows:

5) Assume that $\alpha_1, \dots, \alpha_k$ are the distinct **eigenvalues** of A. Set $\mathcal{B} = \mathcal{B}_{\alpha_1} \cup \mathcal{B}_{\alpha_2} \cup \dots \cup \mathcal{B}_{\alpha_k}$ where the α_i are taken in the same order as they occur in the matrix D. Now let P be the matrix whose sequence of columns is equal to $\mathcal{B}$. This is a **diagonalizing matrix** for A,

Example 7.2.8. Determine if the matrix $A = \begin{pmatrix} -7 & 5 \\ -4 & 2 \end{pmatrix}$ is **diagonalizable**. If it is, find a **basis** of $\mathbb{R}^2$ consisting of **eigenvectors** for A, a **diagonalizing matrix** and a **diagonal form**.

The **characteristic polynomial** is $\chi_A(\lambda) = det\,(A - \lambda I_2) = det\begin{pmatrix} -7-\lambda & 5 \\ -4 & 2-\lambda \end{pmatrix} = (-7-\lambda)(2-\lambda) + 20 = \lambda^2 + 5\lambda + 6 = (\lambda+2)(\lambda+3)$.

Consequently, there are two distinct **eigenvalues** , -2, -3, each with **algebraic multiplicity** one and so this matrix is **diagonalizable** (to $diag\{-2,-3\}$). We find vectors which **span** the **eigenspaces**.

Computing eigenvectors for the eigenvalue $\lambda = -2$:

$A - (-2)I_2 = A + 2I_2 = \begin{pmatrix} -5 & 5 \\ -4 & 4 \end{pmatrix}$. This has **reduced echelon form** $\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$.

Then $E_{-2} = Span\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$.

Computing eigenvectors for the eigenvalue $\lambda = -3$:

$A - (-3)I_2 = A + 3I_2 = \begin{pmatrix} -4 & 5 \\ -4 & 5 \end{pmatrix}$. This has **reduced echelon form** $\begin{pmatrix} 1 & -\frac{5}{4} \\ 0 & 0 \end{pmatrix}$.

From this we deduce that $E_{-3} = Span\left(\begin{pmatrix} \frac{5}{4} \\ 1 \end{pmatrix}\right)$.

Of course, any scalar multiple works just as well as so we multiply by 4 to "clear fractions". So, $E_{-3} = Span\left(\begin{pmatrix} 5 \\ 4 \end{pmatrix}\right)$. $\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \end{pmatrix}\right)$ is a **basis** of **eigenvectors** for A. A **diagonalizing matrix** is $P = \begin{pmatrix} 1 & 5 \\ 1 & 4 \end{pmatrix}$. Then $P^{-1}AP = diag\{-2,-3\}$.

We remark that the matrix $P' = \begin{pmatrix} 5 & 1 \\ 4 & 1 \end{pmatrix}$ will diagonalize A to $diag\{-3,-2\}$.

Example 7.2.9. Determine if the matrix $A = \begin{pmatrix} -8 & 5 \\ -5 & 2 \end{pmatrix}$ is **diagonalizable**. If it is, find a **basis** of **eigenvectors**, a **diagonalizing matrix**, and a **diagonal form**.

The **characteristic polynomial** is $\chi_A(\lambda) = det\,(A - \lambda I_2) = det\begin{pmatrix} -8-\lambda & 5 \\ -5 & 2-\lambda \end{pmatrix} = (-8-\lambda)(2-\lambda) + 25 = \lambda^2 + 6\lambda - 16 + 25 = \lambda^2 + 6\lambda + 9 = (\lambda+3)^2$.

Thus, there is one **eigenvalue**, -3, with **algebraic multiplicity** two. We find a **basis** for the **eigenspace**:

$A-(-3)I_2 = A+3I_2 = \begin{pmatrix} -5 & 5 \\ -5 & 5 \end{pmatrix}$. This has **reduced echelon form** $\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$. This matrix has **rank** one and **nullity** one and therefore the **geometric multiplicity** of the **eigenvalue** -3 is one and this matrix is not **diagonalizable**. The vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a **basis** for the **eigenspace** E_{-3}.

Example 7.2.10. Determine if the matrix $A = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 0 & 1 \\ 6 & -4 & 4 \end{pmatrix}$ is **diagonalizable**. If it is, find a **basis** of **eigenvectors**, a **diagonalizing matrix**, and a **diagonal form**.

The **characteristic polynomial** is $\chi_A(\lambda) =$

$$det(A-\lambda I_3) = det \begin{pmatrix} 2-\lambda & 0 & 0 \\ 3 & -\lambda & 1 \\ 6 & -4 & 4-\lambda \end{pmatrix} =$$

$$(2-\lambda) \times det \begin{pmatrix} -\lambda & 1 \\ -4 & 4-\lambda \end{pmatrix} =$$

$$(2-\lambda)[(-\lambda)(4-\lambda)+4] = (2-\lambda)(\lambda^2-4\lambda+4) = -(\lambda-2)^3.$$

So there is one **eigenvalue**, 2, with **algebraic multiplicity** three. We find a **basis** for the **eigenspace** and see if the **geometric multiplicity** is three and determine whether or not it is **diagonalizable**:

$A-2I_3 = \begin{pmatrix} 0 & 0 & 0 \\ 3 & -2 & 1 \\ 6 & -4 & 2 \end{pmatrix}$ This matrix has **reduced echelon form** $R = \begin{pmatrix} 1 & -\frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

At this point, since the **rank** of this matrix is one, we know its **nullity** is two and therefore the **geometric multiplicity** of 2 is two and the matrix is not **diagonalizable**. We nonetheless find a **basis** for the **eigenspace** to once again demonstrate how to do this.

The matrix R is the **coefficient matrix** of the **homogeneous linear system**

$$x - \frac{2}{3}y + \frac{1}{3}z = 0.$$

The variables y, z are **free variables**. Setting $y = s, z = t$ we get the following general solution:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{2}{3}s - \frac{1}{3}t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} \frac{2}{3} \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -\frac{1}{3} \\ 0 \\ 1 \end{pmatrix}. \text{ Thus, } E_2 = Span\left(\begin{pmatrix} \frac{2}{3} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{3} \\ 0 \\ 1 \end{pmatrix}\right).$$

For those who don't like fractions, we can rescale the **basis** vectors (in both cases by 3) and get $E_2 = Span\left(\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}\right)$.

Example 7.2.11. Determine if the matrix $A = \begin{pmatrix} 5 & 1 & 3 \\ 5 & 5 & 7 \\ -4 & -2 & -3 \end{pmatrix}$ is **diagonalizable**. If it is, find a **basis** of **eigenvectors**, a **diagonalizing matrix**, and a **diagonal form**.

The **characteristic polynomial** is $\chi_A(\lambda) =$

$$det\ (A - \lambda I_3) = det \begin{pmatrix} 5-\lambda & 1 & 3 \\ 5 & 5-\lambda & 7 \\ -4 & -2 & -3-\lambda \end{pmatrix} =$$

$$(5-\lambda)(5-\lambda)(-3-\lambda) - 30 - 28 + 12(5-\lambda) - 5(-3-\lambda) + 14(5-\lambda) =$$

$$-\lambda^3 + 7\lambda^2 + 5\lambda - 75 - 30 - 28 + 60 - 12\lambda + 15 + 5\lambda + 70 - 14\lambda =$$

$$-\lambda^3 + 7\lambda^2 - 16\lambda + 12 = -(\lambda^3 - 7\lambda^2 + 16\lambda - 12).$$

We check for integer roots and so use divisors of 12 (the constant term). After some experimentation find that 2 and 3 are roots. Factoring these out that we get

$$\lambda^3 - 7\lambda^2 + 16\lambda - 12 = (\lambda-2)^2(\lambda-3)$$

Thus, there are two distinct **eigenvalues**: 2 with **algebraic multiplicity** two and 3 with algebraic multiplicity one. We now get bases for the **eigenspaces** and compute the **geometric multiplicities**.

Computing a basis for the 2-eigenspace :

$A - 2I_3 = \begin{pmatrix} 3 & 1 & 3 \\ 5 & 3 & 7 \\ -4 & -2 & -5 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{3}{2} \\ 0 & 0 & 0 \end{pmatrix}$.

Since its **rank** is two, the **nullity** is 1 and so $dim(E_2) = 1$. Thus, the **geometric multip-**

licity of 2 is one and this matrix is not **diagonalizable**. We remark that the vector $\begin{pmatrix} -1 \\ -3 \\ 2 \end{pmatrix}$ is an **eigenvector** for 2. By the procedure described in **Method** (7.1.3) we find that the vector $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$ is an **eigenvector** for the **eigenvalue** 3.

Example 7.2.12. Determine if the matrix $A = \begin{pmatrix} 4 & -2 & 1 \\ 4 & -2 & 2 \\ 2 & -2 & 3 \end{pmatrix}$ is **diagonalizable**. If it is, find a **basis** of **eigenvectors**, a **diagonalizing matrix**, and a **diagonal form**.

The **characteristic polynomial** is $\chi_A(\lambda) =$

$$det\ (A - \lambda I_3) = \begin{pmatrix} 4-\lambda & -2 & 1 \\ 4 & -2-\lambda & 2 \\ 2 & -2 & 3-\lambda \end{pmatrix} =$$

$$(4-\lambda)(-2-\lambda)(3-\lambda) - 8 - 8 - 2(-2-\lambda) + 4(4-\lambda) + 8(3-\lambda) =$$

$$-\lambda^3 + 5\lambda^2 + 2\lambda - 24 - 8 - 8 + 4 + 2\lambda + 16 - 4\lambda + 24 - 8\lambda =$$

$$-\lambda^3 + 5\lambda^2 - 8\lambda + 4 = -(\lambda^3 - 5\lambda^2 + 8\lambda - 4).$$

Trying factors of 4 we find that 1 and 2 are roots and after factoring these out that

$$\lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 1)(\lambda - 2)^2.$$

Thus, there are two distinct **eigenvalues**: 2 with **algebraic multiplicity** two and 1 with algebraic multiplicity one. We now get bases for the **eigenspaces** and compute the **geometric multiplicities**.

Computing a basis for the 2-eigenspace :

$A - 2I_3 = \begin{pmatrix} 2 & -2 & 1 \\ 4 & -4 & 2 \\ 2 & -2 & 1 \end{pmatrix}$. The **reduced echelon form** of this matrix is $S = \begin{pmatrix} 1 & -1 & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.
Since the **rank** is one the **nullity** is two and so the **geometric multiplicity** of the **eigenvalue** 2 is two and the matrix is **diagonalizable matrix**. We get a **basis** for E_2 :

The **homogeneous linear system** with **coefficient matrix** equal to S is $x - y + \frac{1}{2}z = 0$. Setting $y = s, z = t$ we get the general solution

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} s - \frac{1}{2}t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -\frac{1}{2}t \\ 0 \\ t \end{pmatrix}.$$

Then $E_2 = Span\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{2} \\ 0 \\ 1 \end{pmatrix}\right) = Span\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}\right).$

The second **basis** is obtained by scaling the second vector by a factor of two to "clear fractions."

Computing a basis for the 1-eigenspace :

$A - I_3 = \begin{pmatrix} 3 & -2 & 1 \\ 4 & -3 & 2 \\ 2 & -2 & 2 \end{pmatrix}$. The **reduced echelon form** of this matrix is $R = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$.

The matrix R is the **coefficient matrix** of the **homogeneous linear system**

$$\begin{array}{ccccc} x & & - & z & = & 0 \\ & y & - & 2z & = & 0 \end{array}.$$

Setting the **free variable** z equal to t we get the general solution

$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} t \\ 2t \\ t \end{pmatrix} = t\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$. So, $E_1 = Span\left(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}\right)$.

The matrix $\begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$ is a **diagonalizing matrix** and will diagonalize A to $diag\{1, 2, 2\}$.

There are other **diagonalizing matrices** that we can obtain from these **eigenvectors**.

For example, the matrix $\begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 2 & 1 \end{pmatrix}$ is a **diagonalizing matrix** and diagonalizes A to $diag\{2, 2, 1\}$.

Example 7.2.13. Determine if the matrix $A = \begin{pmatrix} 2 & -2 & -2 \\ -2 & 5 & 1 \\ 8 & -16 & -6 \end{pmatrix}$ is **diagonalizable**.

If it is, find a **basis** of **eigenvectors**, a **diagonalizing matrix**, and a **diagonal form**.

The **characteristic polynomial** is $\chi_A(\lambda) =$

$$det\ (A - \lambda I_3) = \begin{pmatrix} 2-\lambda & -2 & -2 \\ -2 & 5-\lambda & 1 \\ 8 & -16 & -6-\lambda \end{pmatrix} =$$

$$(2-\lambda)(5-\lambda)(-6-\lambda) - 64 - 16 + 16(5-\lambda) + 16(2-\lambda) - 4(-6-\lambda) =$$

$$-\lambda^3 + \lambda^2 + 32\lambda - 60 - 64 - 16 + 80 - 16\lambda + 32 - 16\lambda + 24 + 4\lambda =$$

$$-\lambda^3 + \lambda^2 + 4\lambda - 4 = -(\lambda^3 - \lambda^2 - 4\lambda + 4).$$

From inspection we immediately see that 1 is a root. Factoring out $(\lambda - 1)$ we get $-(\lambda^2 - 4)$ and so there are three distinct eigenvalues -2, 1, 2 each with **algebraic multiplicity** one. Consequently, at this point, we know that A is a **diagonalizable matrix**. We continue and find an **eigenvector** for each **eigenvalue** and thus a **basis** of **eigenvalues** and a **diagonalizing matrix**.

Finding a basis for the (-2)-eigenspace:

$A - (-2)I_3 = A + 2I_3 = \begin{pmatrix} 4 & -2 & -2 \\ -2 & 7 & 1 \\ 8 & -16 & -4 \end{pmatrix}$. The **reduced echelon form** is $\begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

The **null space** is spanned by $\begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$ or, after scaling by 2, $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$. So $E_{-2} = Span\left(\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}\right)$.

Finding a basis for the 1-eigenspace:

$A - I_3 = \begin{pmatrix} 1 & -2 & -2 \\ -2 & 4 & 1 \\ 8 & -16 & -7 \end{pmatrix}$. The **reduced echelon form** is $\begin{pmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. The **null space** is spanned by $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$ and so $E_1 = Span\left(\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}\right)$.

Finding a basis for the 2-eigenspace:

$A - 2I_3 = \begin{pmatrix} 0 & -2 & -2 \\ -2 & 3 & 1 \\ 8 & -16 & -8 \end{pmatrix}$. The **reduced echelon form** is $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Thus, $E_2 = Span\left(\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}\right)$. We have the following **basis** of **eigenvectors**:

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

The matrix $\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -1 \\ 2 & 0 & 1 \end{pmatrix}$ is a **diagonalizing matrix**.

Example 7.2.14. As a final example, consider the matrix $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} = 3I_3$. The **characteristic polynomial** of this matrix is $(3-\lambda)^3$. So 3 is an **eigenvector** with **algebraic multiplicity** three. Since $A - 3I_3 = \mathbf{0}_{3\times 3}$ any **basis** is a basis of **eigenvectors** (hey, the matrix was already diagonal). In one of the **challenge exercises** you are asked to prove if A is $n \times n$ matrix with an **eigenvalue** α with **algebraic multiplicity** n then A is **diagonalizable matrix** if and only if $A = \alpha I_n$.

Exercises

In each of the following

a) Determine the **eigenvalues** of the matrix and their **algebraic multiplicity**;
b) Find a **basis** for each **eigenspace** and determine the **geometric multiplicity**;
c) State whether the matrix is **diagonalizable**;
d) If it is a **diagonalizable matrix**, write down a **basis** of **eigenvectors** and a **diagonalizing matrix**. See **Method** (7.2.1).

1. $\begin{pmatrix} 7 & 6 \\ -6 & -5 \end{pmatrix}$

2. $\begin{pmatrix} -4 & -3 \\ 3 & 2 \end{pmatrix}$

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

4. $\begin{pmatrix} -5 & -7 \\ 1 & 3 \end{pmatrix}$

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

6. $\begin{pmatrix} -5 & -4 & 3 \\ -1 & 4 & -1 \\ -7 & 4 & 1 \end{pmatrix}$

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

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

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

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

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

12. $\begin{pmatrix} 1 & 3 & 6 \\ 6 & 4 & 12 \\ -3 & -3 & -8 \end{pmatrix}$

13. $\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$

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

15. $\begin{pmatrix} -1 & 3 & -3 \\ 0 & 4 & -6 \\ 0 & 3 & -4 \end{pmatrix}$

16. $\begin{pmatrix} -1 & -1 & -3 \\ 3 & 3 & 3 \\ -5 & -5 & -3 \end{pmatrix}$

17. $\begin{pmatrix} 4 & 2 & 5 \\ -3 & -1 & -5 \\ -2 & -2 & -3 \end{pmatrix}$

In exercises 18 - 22 answer true or false and give an explanation.

18. If the square matrix A is **diagonalizable** then every **eigenvalue** has **algebraic multiplicity** one.

19. If Q is an **invertible** $n \times n$ matrix then there exists a **diagonalizable matrix** A such that $Q^{-1}AQ$ is a diagonal matrix.

20. The only 2×2 **diagonalizable matrix** which has **eigenvalue** -1 with **algebraic multiplicity** two is $-I_2$.

21. If A is an $n \times n$ **diagonalizable matrix** then A is **invertible**.

22. If A is an $n \times n$ **diagonalizable matrix** then A^{Tr} is **diagonalizable**.

Challenge Exercises (Problems)

Let A be an $n \times n$ matrix. A **subspace** U of $\mathbb{R}^n$ is said to be $A-$invariant if $\boldsymbol{u} \in U$ implies that $A\boldsymbol{u} \in U$.

1. Let A be an $n \times n$ **diagonalizable matrix** and U an $A-$invariant **subspace** of $\mathbb{R}^n$. Prove that U contains an **eigenvector** for A.

2. Prove that U has a basis consisting of eigenvectors for A.

3. Assume that A and B commute and are $n \times n$ **diagonalizable matrices**. Prove that A and B have a common **eigenvector** , i.e. a vector $\boldsymbol{v} \neq \boldsymbol{0}_n$ such that $A\boldsymbol{v} = \alpha\boldsymbol{v}, B\boldsymbol{v} = \beta\boldsymbol{v}$ for scalars α, β.

4. Let A be an $n \times n$ matrix with a unique **eigenvalue** α of **algebraic multiplicity** n. Prove that A is **diagonalizable** if and only if $A = \alpha I_n$.

Let A be an $n \times n$ matrix and U an $A-$invariant **subspace** of $\mathbb{R}^n$. U is said to have an $A-$complement if there is an $A-$invariant subspace W such that $U \cap W = \{0\}, U + W = \mathbb{R}^n$.

5. Let A be an $n \times n$ **diagonalizable matrix**. Prove that every $A-$invariant **subspace** has an $A-$complement.

6. Prove if A is a **diagonalizable matrix** then A^{Tr} is also diagonalizable.

7. Let J_n be the $n \times n$ matrix all of whose entries are 1's. Prove that J_n is a **diagonalizable matrix** and display a **basis** of $\mathbb{R}^n$ consisting of **eigenvectors** for J_n.

8. For real numbers a, b prove that $aJ_n + bI_n$ is a **diagonalizable matrix** and find a **basis** of $\mathbb{R}^n$ consisting of **eigenvectors** for $aJ_n + bI_n$.

Quiz Solutions

1. $\chi_A(\lambda) = det\ (A - \lambda I_2) = det\ \begin{pmatrix} 5-\lambda & 4 \\ -4 & -3-\lambda \end{pmatrix} =$

$$(5-\lambda)(-3-\lambda) + 16 = \lambda^2 - 2\lambda - 15 + 16 = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2.$$

There is a unique **eigenvalue**, 1, with **algebraic multiplicity** two. We find a **basis** for the **eigenspace**.

$A - I_2 = \begin{pmatrix} 4 & 4 \\ -4 & -4 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. $E_1 = null(A - I_2) = Span\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)$. The **geometric multiplicity** of the **eigenvalue** 1 is one. Not right, see **Method** (7.1.1), **Method** (7.1.2), **Method** (7.1.3).

2. $\chi_A(\lambda) = det\ (A - \lambda I_2) = det\ \begin{pmatrix} 1-\lambda & 4 \\ 4 & 1-\lambda \end{pmatrix} =$

$$(1-\lambda)(1-\lambda) - 16 = \lambda^2 - 2\lambda + 1 - 16 = \lambda^2 - 2\lambda - 15 = (\lambda - 5)(\lambda + 3).$$

There are two **eigenvalues**, -3 and 5, each with **algebraic multiplicity** one. We find **basis** for the each of the **eigenspaces** E_{-3} and E_5.

Finding a basis for the (-3)-eigenspace:

$A - (-3)I_2 = A + 3I_2 = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}$. The **reduced echelon form** is $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. Then $E_{-3} = Span\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)$.

Finding a basis for the 5-eigenspace:

$A - 5I_2 = \begin{pmatrix} -4 & 4 \\ 4 & -4 \end{pmatrix}$. The **reduced echelon form** is $\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$. Thus, $E_5 = Span\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$.

Each **eigenvalue** has **geometric multiplicity** one. Not right, see **Method** (7.1.1), **Method** (7.1.2), **Method** (7.1.3).

3. $\chi_A(\lambda) = det\ (A - \lambda I_3) = det\ \begin{pmatrix} 3-\lambda & 2 & -2 \\ 5 & 3-\lambda & -5 \\ 5 & 2 & -4-\lambda \end{pmatrix} =$

$$(3-\lambda)(3-\lambda)(-4-\lambda) - 20 - 50 + 10(3-\lambda) + 10(3-\lambda) - 10(-4-\lambda) =$$

$$-\lambda^3 + 2\lambda^2 + 15\lambda - 36 - 20 - 50 + 30 - 10\lambda + 30 - 10\lambda + 40 + 10\lambda =$$

$$-\lambda^3 + 2\lambda^2 + 5\lambda - 6 = -(\lambda^3 - 2\lambda^2 - 5\lambda + 6) = -(\lambda+2)(\lambda-1)(\lambda-3)$$

There are three distinct **eigenvalues**, -2, 1, 3, each with **algebraic multiplicity** one. We find a **basis** for each of the **eigenspaces**.

Finding a basis of the (-2)-eigenspace:

$A - (-2)I_3 = A + 2I_3 = \begin{pmatrix} 5 & 2 & -2 \\ 5 & 5 & -5 \\ 5 & 2 & -2 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. Then $E_{-2} = Span\left(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right)$.

Finding a basis of the 1-eigenspace:

$A - I_3 = \begin{pmatrix} 2 & 2 & -2 \\ 5 & 2 & -5 \\ 5 & 2 & -5 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Then $E_1 = Span\left(\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\right)$.

Finding a basis of the 3-eigenspace:

$A - 3I_3 = \begin{pmatrix} 0 & 2 & -2 \\ 5 & 0 & -5 \\ 5 & 2 & -7 \end{pmatrix}$. The reduced echelon form is $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. Then $E_3 = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right)$.

Each **eigenvalue** has **geometric multiplicity** one. Not right, see **Method** (7.1.1), **Method** (7.1.2), **Method** (7.1.3).

7.3. Complex Eigenvalues of Real Matrices

In this section we consider a real $n \times n$ matrix A with complex eigenvalues.

Am I Ready for This Material

The following are those concepts which are used extensively in this section:

null space

diagonal matrix

linearly independent sequence

basis of $\mathbb{R}^n$

change of basis matrix

determinant of an n x n matrix

similar matrices

eigenvector of an n x n matrix

eigenvalue of an n x n matrix

characteristic polynomial of an n x n matrix

characteristic equation of an n x n matrix

algebraic multiplicity of an eigenvalue of an n x n matrix

eigenspace of a matrix for an eigenvalue

geometric multiplicity of an eigenvalue of an n x n matrix

diagonalizable marix

diagonal form

diagonalizing matrix

complex field

vector space over a field

complex vector space

Quiz

1. a) Determine the **characteristic polynomial** of the matrix $\begin{pmatrix} 2 & 5 \\ -1 & 4 \end{pmatrix}$ and extract all its roots (real and non-real complex).

b) Perform the required multiplication: $\begin{pmatrix} 2 & 5 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 1-2i \\ 1 \end{pmatrix}$.

c) Perform the required scalar multiplication: $(3+2i)\begin{pmatrix} 1-2i \\ 1 \end{pmatrix}$

2. Find the **null space** of the complex matrix $\begin{pmatrix} 2-3i & 5-i \\ 1-i & 2 \end{pmatrix}$.

3. Determine the **characteristic polynomial** of the matrix $\begin{pmatrix} 4 & -2 & -4 \\ 0 & 1 & -1 \\ 2 & -1 & -1 \end{pmatrix}$ and extract all its roots (real and non-real complex).

4. Find a **basis** in $\mathbb{C}^3$ for the **null space** of the complex matrix $\begin{pmatrix} 3+i & -2 & -4 \\ 0 & i & -1 \\ 2 & -1 & -2+i \end{pmatrix}$

Quiz Solutions

New Concepts

In this section we introduce a single new concept:

invariant subspace of a matrix A

Theory (Why It Works)

As mentioned in **Remark** (7.8) all the definitions introduced in Section (7.1) and Section (7.2) apply. For example,

Definition 7.8. Let A be a complex $n \times n$ matrix. A vector $\boldsymbol{z} \in \mathbb{C}^n$ is an ***eigenvector*** of A with ***eigenvalue*** $\alpha \in \mathbb{C}$ if $A\boldsymbol{z} = \alpha\boldsymbol{z}$.

We can define the ***algebraic multiplicity***, the ***geometric multiplicity*** of a complex eigenvalue, the notion of being ***diagonalizable*** (to a complex diagonal matrix) and so on.

It is not our intent in this section to duplicate the results of Section (7.1) and Section (7.2), that is unnecessary: all the proofs carry over with little change in detail. Rather in this section we study real $n \times n$ matrices with non-real complex eigenvalues. Such a matrix is not **diagonalizable matrix** via a real matrix but may be **similar** to a complex **diagonal matrix** (via a complex **diagonalizing matrix**). The criterion for this is the same as the case when all the **eigenvalues** are real: there must exist a **basis** for $\mathbb{C}^n$ consisting of **eigenvectors** for A. It is our purpose here to uncover what the existence of complex **eigenvalues** reveals about the matrix as a real matrix. Before proceeding to an example we make an observation:

Remark 7.9. Suppose A is a real $n \times n$ matrix and $\boldsymbol{z} \in \mathbb{C}^n$ is a complex eigenvector for a non-real, **complex eigenvalue** $\alpha = a + bi$. Then $Span(\boldsymbol{z}) \cap \mathbb{R}^n$ is just the **zero vector**, $\mathbf{0}_n$. In particular, $\boldsymbol{z}$ is non-real.

Note that by $Span(\boldsymbol{z})$ we mean the complex span, all multiples of $\boldsymbol{z}$ by a complex number. Suppose to the contrary that $\boldsymbol{w} \in Span(\boldsymbol{z}) \cap \mathbb{R}^n, \boldsymbol{w} \neq \mathbf{0}_n$. Then $A\boldsymbol{w} \in \mathbb{R}^n$. However, as a non-zero multiple of $\boldsymbol{z}, \boldsymbol{w}$ is a complex eigenvector of A for the non-real complex eigenvalue $\alpha \notin \mathbb{R}$. Then $A\boldsymbol{w} = \alpha\boldsymbol{w}$ does not belong to $\mathbb{R}^n$.

Example 7.3.1. Let θ be real number, $0 \leq \theta \leq 2\pi$. Recall we defined the transformation $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates a vector in the plane by the angle θ in the counterclockwise direction. We demonstrated in **Example** (3.1.4) that this is a **matrix transformation** with **standard matrix** $A = \begin{pmatrix} cos\ \theta & -sin\ \theta \\ sin\ \theta & cos\ \theta \end{pmatrix}$.

The **characteristic polynomial** of A is

$$\chi_A(\lambda) = det(A - \lambda I_2) = det \begin{pmatrix} cos\ \theta - \lambda & -sin\ \theta \\ sin\ \theta & cos\ \theta\lambda \end{pmatrix} =$$

$$(cos\ \theta - \lambda)(cos\ \theta) - (-sin\ \theta)(sin\ \theta) = \lambda^2 - 2cos\ \theta + 1 \tag{7.4}$$

The roots of $\chi_A(\lambda)$ are

$$\frac{2cos\ \theta \pm \sqrt{(2cos\ \theta)^2 - 4}}{2} = \frac{2cos\ \theta \pm \sqrt{(4cos^2\ \theta - 4}}{2} =$$

$$\frac{2cos\ \theta \pm 2\sqrt{(1 - cos^2\ \theta}\sqrt{-1}}{2} = cos\ \theta \pm i\ sin\ \theta \tag{7.5}$$

So roots of the **characteristic polynomial** are the pair of **conjugate complex numbers** $cos\ \theta + i\ sin\ \theta$ and $cos\ \theta - i\ sin\ \theta$.

We claim that the complex vector $\begin{pmatrix} i \\ 1 \end{pmatrix}$ is a **eigenvector** with **eigenvalue** $cos\ \theta + i\ sin\ \theta$ and the vector $\begin{pmatrix} -i \\ 1 \end{pmatrix}$ is a eigenvector with eigenvalue $cos\ \theta - i\ sin\ \theta$. We check:

$$\begin{pmatrix} cos\ \theta & -sin\ \theta \\ sin\ \theta & cos\ \theta \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} i\ cos\theta - sin\ \theta \\ i\ sin\ \theta + cos\ \theta \end{pmatrix} = (cos\ \theta + i\ sin\ \theta) \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

$$\begin{pmatrix} cos\ \theta & -sin\ \theta \\ sin\ \theta & cos\ \theta \end{pmatrix} \begin{pmatrix} -i \\ 1 \end{pmatrix} = \begin{pmatrix} -i\ cos\theta - sin\ \theta \\ -i\ sin\ \theta + cos\ \theta \end{pmatrix} = (cos\ \theta - i\ sin\ \theta) \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Example 7.3.2. Consider the matrix $A = \begin{pmatrix} 1 & 5 \\ -1 & -3 \end{pmatrix}$. The **characteristic polynomial** of this matrix is $\chi_A(\lambda) =$

$$det \begin{pmatrix} 1-\lambda & 5 \\ -1 & -3-\lambda \end{pmatrix} =$$

$$(1-\lambda)(-3-\lambda) + 5 = \lambda^2 + 2\lambda - 3 + 5 = \lambda^2 + 2\lambda + 2 \tag{7.6}$$

The roots of the **characteristic polynomial** are $-1+i$ and $-1-i$, which is a **conjugate pair of complex numbers**.

By straightforward calculations it is easy to check that the complex vector $\begin{pmatrix} -2-i \\ 1 \end{pmatrix}$ is an **eigenvector** with **eigenvalue** $-1+i$ and the vector $\begin{pmatrix} -2+i \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue $-1-i$.

Notice in Example (7.3.1) and Example (7.3.2) the complex roots of $\chi_A(\lambda)$ occurred as a **conjugate pair of complex numbers**. Also observe that the **eigenvectors** are complex conjugates in the sense that if one **eigenvector** is $\boldsymbol{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ then the other is $\overline{\boldsymbol{z}} = \begin{pmatrix} \overline{z_1} \\ \overline{z_2} \end{pmatrix}$. The first of these phenomena, that the **complex eigenvalues** of a real matrix occur as **conjugate pairs of complex numbers** is a consequence of the following theorem about roots of real polynomials.

Theorem 7.3.1. *Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial with real coefficients: $a_i \in \mathbb{R}$, for $i = 0, 1, \ldots, n$. If $z \in \mathbb{C}$ is a root of $f(x)$ then the complex conjugate, $\overline{z}$, is also a root.*

Proof. If z is a root of $f(x)$ then $a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = 0$. Taking the complex conjugate of both sides we get $\overline{f(z)} = 0$. However, by part 3) of Theorem

(5.6.2), for each $i, \overline{a_i z^i} = \overline{a_i}\ \overline{z^i} = a_i \overline{z}^i$, the latter equality since, for the real number $a_i, \overline{a_i} = a_i$. It then follows that

$$\overline{f(z)} = a_n \overline{z}^n + a_{n-1}\overline{z}^{n-1} + \cdots + a_1\overline{z} + a_0 = f(\overline{z}) \tag{7.7}$$

Since $\overline{f(z)} = 0$ this together with (7.7) implies that $f(\overline{z}) = 0$ and $\overline{z}$ is a root of $f(x)$. □

Before proceeding to our next important theorem we prove a lemma:

Lemma 7.3.2. *Let A be an $m \times n$ complex matrix and B be an $n \times p$ complex matrix. Then $\overline{AB} = \overline{A}\ \overline{B}$.*

Proof. First suppose that $\boldsymbol{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$ and $\boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ are complex vectors. By **Theorem** (5.6.2) we have that the complex conjugate of $\boldsymbol{z} \cdot \boldsymbol{w} = z_1 w_1 + z_2 w_2 + \ldots z_n w_n$ is

$$\overline{z_1 w_1} + \cdots + \overline{z_n w_n} = \overline{z_1}\ \overline{w_1} + + \ldots \overline{z_n}\ \overline{w_n} = \overline{\boldsymbol{z}} \cdot \overline{\boldsymbol{w}}. \tag{7.8}$$

Now let A be an $m \times n$ complex matrix and B be an $n \times p$ complex matrix. Let $\boldsymbol{a}_i^{Tr}$ be the i^{th} row of A and $\boldsymbol{b}_j$ be the j^{th} column of B. Then the $(i,j)-$entry of AB is $\boldsymbol{a}_i \cdot \boldsymbol{b}_j$ and the $(i,j)-$entry of $\overline{AB}$ is $\overline{\boldsymbol{a}_i \cdot \boldsymbol{b}_j}$. On the other hand, the i^{th} row of $\overline{A}$ is $\overline{\boldsymbol{a}_i^{Tr}}$ and the j^{th} column of $\overline{B}$ is $\overline{\boldsymbol{b}_j}$. Therefore the $(i,j)-$entry of $\overline{A}\overline{B}$ is $\overline{\boldsymbol{a}_i} \cdot \overline{\boldsymbol{b}_j}$.

However, by (7.8) $\overline{\boldsymbol{a}_i} \cdot \overline{\boldsymbol{b}_j} = \overline{\boldsymbol{a}_i \cdot \boldsymbol{b}_j}$. Since this holds for each i and j we conclude that $\overline{AB} = \overline{A}\ \overline{B}$.

□

Using Theorem (7.3.1) and Lemma (7.3.2) we can prove that the phenomena observed in **Example** (7.3.1) and **Example** (7.3.2) always occur:

Theorem 7.3.3. *Let A be a real $n \times n$ matrix and and assume the complex number α is a root of the* **characteristic polynomial** *$\chi_A(\lambda) = det(A - \lambda I_n)$. Then the following hold:*

1) The **complex conjugate** *of $\alpha, \overline{\alpha}$ is also a root of $\chi_A(\lambda)$.*

2) If the complex vector $\boldsymbol{z}$ is an **eigenvector** *of A with* **eigenvalue** *α then $\overline{\boldsymbol{z}}$ is an an eigenvector of A with eigenvalue $\overline{\alpha}$.*

Proof. 1) Since $\chi_A(\lambda)$ is a polynomial with real coefficients this is an immediate consequence of **Theorem** (7.3.1) .

2) Suppose $\boldsymbol{z}$ is an eigenvector of A with eigenvalue α. This means that $A\boldsymbol{z} = \alpha\boldsymbol{z}$. Take complex conjugates of both sides. $\overline{A\boldsymbol{z}} = \overline{\alpha\boldsymbol{z}}$.

By Lemma (7.3.2) we have $\overline{A\boldsymbol{z}} = \overline{A}\,\overline{\boldsymbol{z}} = A\overline{\boldsymbol{z}}$ since A is a real matrix.

On the other hand $\overline{\alpha\boldsymbol{z}} = \overline{\alpha}\,\overline{\boldsymbol{z}}$. Thus, $A\overline{\boldsymbol{z}} = \overline{\alpha}\,\overline{\boldsymbol{z}}$ which shows that $\overline{\boldsymbol{z}}$ is an **eigenvector** of A with **eigenvalue** $\overline{\alpha}$.

□

We will soon see how to exploit the last result to show that associated with each **complex eigenvector** $\boldsymbol{z}$ for A with **eigenvalue** α there is a two dimensional **subspace** of $\mathbb{R}^n$ which is left invariant by A (defined immediately below). This will have lead to very strong results when the matrix A is 2×2 or 3×3.

Definition 7.9. Let A be an $n \times n$ real matrix. A **subspace** W of $\mathbb{R}^n$ is said to be $A-$***invariant*** if for every $\boldsymbol{w} \in W$, $A\boldsymbol{w} \in W$.

We will need the following lemma later in the section when we discuss real canonical forms of 3×3 real matrices.

Lemma 7.3.4. *Assume the $n \times n$ matrices A and B commute, that is, $AB = BA$. Then the* **column space** *of B is $A-$invariant.*

Proof. We first remark that a vector $\boldsymbol{y}$ is in **column space** of B if and only if there exists a vector $\boldsymbol{x}$ such that $\boldsymbol{y} = B\boldsymbol{x}$. So, assume $\boldsymbol{y}$ is in the **column space** of B and let $\boldsymbol{x}$ be an **n-vector** such that $\boldsymbol{y} = B\boldsymbol{x}$. We need to show that $A\boldsymbol{y}$ is in the column space of B. Set $\boldsymbol{z} = A\boldsymbol{x}$. Then $B\boldsymbol{z} = B(A\boldsymbol{x}) = (BA)\boldsymbol{x} = (AB)\boldsymbol{x} = A(B\boldsymbol{x}) = A\boldsymbol{y}$. Thus, $B\boldsymbol{z} = A\boldsymbol{y}$ and therefore $A\boldsymbol{y}$ is in the **column space** of B as required. □

The next theorem basically shows how to construct a two dimensional invariant subspace for a real $n \times n$ matrix given a **complex eigenvector**.

Theorem 7.3.5. *Let A be an $n \times n$ real matrix and let $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ ($\boldsymbol{x}, \boldsymbol{y}$ real $n-$vectors) be a* **complex eigenvector** *for A with* **eigenvalue** *$\alpha = a + bi$ (a, b real, $b \neq 0$). Then $A\boldsymbol{x} = a\boldsymbol{x} - b\boldsymbol{y}$, $A\boldsymbol{y} = b\boldsymbol{x} + a\boldsymbol{y}$. In particular, the two dimensional space $Span(\boldsymbol{x}, \boldsymbol{y})$ is invariant for A.*

Proof. Since $\boldsymbol{z}$ is an **eigenvector** with **eigenvalue** α we have $A\boldsymbol{z} = \alpha\boldsymbol{z}$. Then $A\boldsymbol{z} = A(\boldsymbol{x} + i\boldsymbol{y}) = A\boldsymbol{x} + iA\boldsymbol{y}$. On the other hand, $\alpha\boldsymbol{z} = (a + bi)(\boldsymbol{x} + i\boldsymbol{y}) = (a\boldsymbol{x} - b\boldsymbol{y}) + (b\boldsymbol{x} + a\boldsymbol{y})i$. Equating the real and imaginary we get $A\boldsymbol{x} = a\boldsymbol{x} - b\boldsymbol{y}$, $A\boldsymbol{y} = b\boldsymbol{x} + a\boldsymbol{y}$ as asserted.

It follows that $Span(\boldsymbol{x}, \boldsymbol{y})$ is left invariant by A. It remains to show that this is **two dimensional**, that is, $\boldsymbol{x}, \boldsymbol{y}$ are not multiples of one another. Suppose, to the contrary, that $\boldsymbol{y} = c\boldsymbol{x}$ for some scalar (necessarily real since both $\boldsymbol{x}$ and $\boldsymbol{y}$ are real). Then $\boldsymbol{z} = \boldsymbol{x} + i(c\boldsymbol{x}) = (1 + ic)\boldsymbol{x}$. But then $\boldsymbol{x} \in Span(\boldsymbol{z})$ contrary to (7.9). $\square$

Example 7.3.3. Let A be the matrix $\begin{pmatrix} 6 & 5 \\ -5 & -2 \end{pmatrix}$. This matrix has **characteristic polynomial** $\chi_A(\lambda) = \lambda^2 - 4\lambda + 13$ with roots $\alpha = 2 + 3i$ and $\overline{\alpha} = 2 - 3i$. The vector $\boldsymbol{z} = \begin{pmatrix} -4 - 3i \\ 5 \end{pmatrix}$ is an **eigenvector with eigenvalue α** and $\overline{\boldsymbol{z}} = \begin{pmatrix} -4 + 3i \\ 5 \end{pmatrix}$ is an eigenvector with eigenvalue $\overline{\alpha}$.

The vector $\boldsymbol{z} = \begin{pmatrix} -4 \\ 5 \end{pmatrix} + i\begin{pmatrix} -3 \\ 0 \end{pmatrix}$. We multiply these vectors by A :

$$A\boldsymbol{x} = \begin{pmatrix} 6 & 5 \\ -5 & -2 \end{pmatrix}\begin{pmatrix} -4 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 10 \end{pmatrix}, A\boldsymbol{y} = \begin{pmatrix} 6 & 5 \\ -5 & -2 \end{pmatrix}\begin{pmatrix} -3 \\ 0 \end{pmatrix} = \begin{pmatrix} -18 \\ 15 \end{pmatrix}.$$

The **reduced echelon form** of the matrix $\begin{pmatrix} -4 & -3 & | & 1 & -18 \\ 5 & 0 & | & 10 & 15 \end{pmatrix}$ is

$$\begin{pmatrix} 1 & 0 & | & 2 & 3 \\ 0 & 1 & | & -3 & 2 \end{pmatrix} \tag{7.9}$$

We can conclude from (7.9) that

$$\begin{pmatrix} 1 \\ 10 \end{pmatrix} = 2\begin{pmatrix} -4 \\ 5 \end{pmatrix} + (-3)\begin{pmatrix} -3 \\ 0 \end{pmatrix}, \begin{pmatrix} -18 \\ 15 \end{pmatrix} = 3\begin{pmatrix} -4 \\ 5 \end{pmatrix} + 2\begin{pmatrix} -3 \\ 0 \end{pmatrix} \tag{7.10}$$

as to be expected by Theorem (7.3.5).

We can express the equalities of (7.10) in terms of matrix multiplication we have

$$\begin{pmatrix} 6 & 5 \\ -5 & -2 \end{pmatrix}\begin{pmatrix} -4 & -3 \\ 5 & 0 \end{pmatrix} = \begin{pmatrix} -4 & -3 \\ 5 & 0 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} \tag{7.11}$$

Alternatively, (7.11) can be written as

$$\begin{pmatrix} -4 & -3 \\ 5 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 6 & 5 \\ -5 & -2 \end{pmatrix} \begin{pmatrix} -4 & -3 \\ 5 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} \tag{7.12}$$

This occurs in general as the next theorem asserts.

Theorem 7.3.6. *Assume that the* 2×2 *real matrix* A *has* **complex eigenvector** $\boldsymbol{z} = \boldsymbol{x}+i\boldsymbol{y}$ *(*$\boldsymbol{x},\boldsymbol{y}$ *real* $n-$*vectors) with* **eigenvalue** $a+bi$ *(*a,b *real numbers,* $b\neq 0$*). Let* P *be the* 2×2 *matrix with columns* $\boldsymbol{x}$ *and* $\boldsymbol{y}$, $P = (\boldsymbol{x}\ \boldsymbol{y})$. *Then* $P^{-1}AP = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$.

Proof. From **Theorem** (7.3.5) we know that

$$A\boldsymbol{x} = a\boldsymbol{x} - b\boldsymbol{y},\ A\boldsymbol{y} = b\boldsymbol{x} + a\boldsymbol{y} \tag{7.13}$$

The relations (7.13) can be written in terms of matrix multiplication. Let P be the 2×2 matrix $(\boldsymbol{x}\ \boldsymbol{y})$. Then

$$AP = P\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{7.14}$$

As in the proof of **Theorem** (7.3.5) the vectors $\boldsymbol{x},\boldsymbol{y}$ are not multiples of one another. Therefore, the matrix P is **invertible** .

Multiplying both sides of (7.14) by P^{-1} on the left hand side yields the desired result. $\square$

With Theorem (7.3.6) we can essentially "classify" all the 2×2 matrices, at least in the sense of writing down a short list of matrices which have the property that every 2×2 matrix is **similar** to one and only one on the list. Before stating and proving this we introduce a related definition.

Recall that two $n\times n$ matrices A and B are **similar** if there is an **invertible matrix** P such $B = P^{-1}AP$.

Definition 7.10. Let A be an $n\times n$ real matrix. The collection of matrices which are **similar** to A is called the ***similarity class of*** A. Any element of the similarity class of A is called a ***representative of the similarity class***.

Theorem 7.3.7. *The following matrices are representatives for all the similarity classes of 2×2 real matrices:*

$$\begin{pmatrix} b & 0 \\ 0 & c \end{pmatrix} \tag{7.15}$$

for b, c real numbers, not necessarily distinct;

$$\begin{pmatrix} b & 0 \\ 1 & b \end{pmatrix} \tag{7.16}$$

for b a real number;

$$\begin{pmatrix} b & c \\ -c & b \end{pmatrix} \tag{7.17}$$

for real numbers b and c.

Proof. There are two major cases: i) the **characteristic polynomial** of A has two real roots b and c (not necessarily distinct); and ii) the roots of $\chi_A(\lambda)$ form a pair of **complex conjugates** $\alpha = b + ci$ and $\overline{\alpha} = b - ci$. In case ii) we have seen that A is **similar** to $\begin{pmatrix} b & c \\ -c & b \end{pmatrix}$ by Theorem (7.3.6). So we may assume that ii) holds.

If b and c are distinct then by **Corollary** (7.2.6), A is **diagonalizable** and **similar** to $\begin{pmatrix} b & 0 \\ 0 & c \end{pmatrix}$. So we may assume that b is an **eigenvalue** of A with **algebraic multiplicity** two. If the **geometric multiplicity** of b is also two then $A = bI_2$ and there is nothing further to show.

Finally, assume that b occurs as an **eigenvalue** with **algebraic multiplicity** two and **geometric multiplicity** one. Let $\boldsymbol{w}$ be an **eigenvector** of A with **eigenvalue** b and let $\boldsymbol{v}$ be any vector in $\mathbb{R}^2$ which is not in $Span(\boldsymbol{w})$, that is, not a multiple of $\boldsymbol{w}$. Since the **geometric multiplicity** of b is one, $\boldsymbol{v}$ is not an **eigenvector** . Moreover, the sequence $(\boldsymbol{v}, \boldsymbol{w})$ is **linearly independent** and a **basis** for $\mathbb{R}^2$.

Since $(\boldsymbol{v}, \boldsymbol{w})$ is a **basis** for $\mathbb{R}^2$ it is then the case that $A\boldsymbol{v}$ is a **linear combination** of $(\boldsymbol{v}, \boldsymbol{w})$, say $A\boldsymbol{v} = d\boldsymbol{v} + e\boldsymbol{w}$. Since $\boldsymbol{v}$ is not an eigenvector $e \neq 0$. Now set $\boldsymbol{v}_1 = \boldsymbol{v}$ and $\boldsymbol{v}_2 = e\boldsymbol{w}$. Then $\boldsymbol{v}_2$ is also an **eigenvector** of A with **eigenvalue** b and $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is a **basis** for $\mathbb{R}^2$ which we denote by $\mathcal{B}$. We now have

$$A\boldsymbol{v}_1 = d\boldsymbol{v}_1 + \boldsymbol{v}_2, A\boldsymbol{v}_2 = b\boldsymbol{v}_2 \tag{7.18}$$

Let $T:\mathbb{R}^2 \to \mathbb{R}^2$ be the **linear transformation** which has **standard matrix** A. Then (7.18) says that the **matrix of T with respect to $\mathcal{B}$** is $\begin{pmatrix} d & 0 \\ 1 & b \end{pmatrix}$.

Set $P = (\boldsymbol{v}_1\ \boldsymbol{v}_2) = P_{\mathcal{B}\to\mathcal{S}}$, where $\mathcal{S}$ is the **standard basis** of $\mathbb{R}^2$. Then by **Theorem** (6.3.2), $P^{-1}AP = \begin{pmatrix} d & 0 \\ 1 & b \end{pmatrix} = B$ and therefore the matrix A is **similar** to the matrix B.

However, **similar matrices** have the same **characteristic polynomial**. By assumption the **characteristic polynomial** of A is $(\lambda - b)^2$ while the **characteristic polynomial** of B is $(\lambda - d)(\lambda - b)$ and therefore $d = b$. This completes the proof. $\square$

Definition 7.11. The matrices in (7.15), (7.16), and (7.17) are referred to as the ***canonical forms for real*** 2×2 ***matrices***

Linear Transformations of $\mathbb{R}^2$ With Complex Eigenvalues

Let $T:\mathbb{R}^2 \to \mathbb{R}^2$ be a **linear transformation** and assume that the **eigenvalues** of T are a conjugate pair of complex numbers, $\alpha = a + bi$ and $\overline{\alpha} = a - bi$. Let A be the **standard matrix** of T (and so the **matrix of T** with respect to the standard basis $\mathcal{S}$). If $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ is an **eigenvector** for A with **eigenvalue** α, where $\boldsymbol{x}, \boldsymbol{y}$ are real vectors, then **Theorem** (7.3.6) implies that the **matrix of T** with respect to $\mathcal{B} = (\boldsymbol{x}, \boldsymbol{y})$ is $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. This latter matrix is "nearly" the matrix of a rotation of $\mathbb{R}^2$ - it is a rotation followed by an expansion or contraction of vectors by a constant factor as we show:

Set $r = \sqrt{a^2+b^2}, x = \frac{a}{r}, y = \frac{b}{r}$. Since $x^2 + y^2 = 1$ there is a unique $\theta, 0 \le \theta < 2\pi$ such that $cos\ \theta = x, sin\ \theta = y$.

Now observe that the matrix $\begin{pmatrix} a & b \\ -b & a \end{pmatrix} =$

$$\begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}\begin{pmatrix} x & y \\ -y & x \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}\begin{pmatrix} cos\ \theta & sin\ \theta \\ -sin\ \theta & cos\ \theta \end{pmatrix}.$$

The matrix $\begin{pmatrix} cos\ \theta & sin\ \theta \\ -sin\ \theta & cos\ \theta \end{pmatrix}$ is the **standard matrix** for the rotation of $\mathbb{R}^2$ by θ in the counterclockwise direction and the matrix $\begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}$ is the matrix of an expansion (if $r > 1$) or a contraction (if $r < 1$).

Therefore, in general, a **linear transformation** of $\mathbb{R}^2$ which has **complex eigenvalues** can be viewed as a rotation followed by a contraction or expansion in a different coordinate system.

Similarity Classes For 3×3 Real Matrices

Suppose now that A is a 3×3 real matrix. Now, the **characteristic polynomial** of A has degree three. If it has a non-real complex root then these come in a conjugate pair $\alpha = a + bi$ $(b \neq 0)$ and $\overline{\alpha} = a - bi$ by **Theorem** (7.3.1) and therefore the **characteristic polynomial** has one real root, c. Let $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ be an **eigenvector** for A with **eigenvalue** α where $\boldsymbol{x}, \boldsymbol{y}$ are real and let $\boldsymbol{w}$ be an **eigenvector** of A with **eigenvalue** c. We then have the following:

$$A\boldsymbol{w} = c\boldsymbol{w}, A\boldsymbol{x} = a\boldsymbol{x} - b\boldsymbol{y}, A\boldsymbol{y} = b\boldsymbol{x} + a\boldsymbol{y} \tag{7.19}$$

Set $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$. Then P is **invertible** and

$$P^{-1}AP = \begin{pmatrix} c & 0 & 0 \\ 0 & a & b \\ 0 & -b & a \end{pmatrix} \tag{7.20}$$

We have therefore proved

Theorem 7.3.8. *Let A be a real 3×3 matrix and assume that A has a non-real. **complex eigenvalue**. Then A is **similar** via a real matrix to a matrix of the form* $\begin{pmatrix} c & 0 & 0 \\ 0 & a & b \\ 0 & -b & a \end{pmatrix}$.

We can also write down **representatives** for all the possible **similarity classes** for a real 3×3 matrices:

Theorem 7.3.9. *Let A be a real 3×3 matrix. Then there is a real* **invertible matrix** *P such that $P^{-1}AP$ is one of the following:*

$$\begin{pmatrix} c & 0 & 0 \\ 0 & a & b \\ 0 & -b & a \end{pmatrix} \tag{7.21}$$

(a, b, c not necessarily distinct)

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \tag{7.22}$$

(a, b, c not necessarily distinct)

$$\begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 0 & b \end{pmatrix} \tag{7.23}$$

(a and b not necessarily distinct)

$$\begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 1 & a \end{pmatrix} \tag{7.24}$$

Proof. The proof divides into several cases: 1) The **characteristic polynomial** of A has one real root (and two conjugate, non-real, complex roots); and 2) the characteristic polynomial of A has three real roots. In turn, 2) has two cases: a) A is **diagonalizable** (for each eigenvalue the **algebraic multiplicity** is equal to its **geometric multiplicity** ; and b) A is **not diagonalizable** (so A has a real **eigenvalue** with **geometric multiplicity** less than its **algebraic multiplicity**). This case, also divides into two cases: i) A has two distinct **eigenvalues** ; and ii) A has a unique **eigenvalue** a with **algebraic multiplicity** three. Finally, this case divides into two cases: α) the **geometric multiplicity** of a is two; and β) the **geometric multiplicity** of a is one.

1) In this case, by Theorem (7.3.8) A is **similar** to a matrix of the form shown in (7.21).

2a) This means that A is **similar** to a matrix of the form shown in (7.22).

The remaining cases are somewhat complicated. We prove here the most difficult case, 2biiβ) to illustrate the methods and the remaining proofs are assigned as **challenge exercises**. There are multiple steps and several general ideas about doing proofs that are present in this one.

2biiβ) Since the **eigenvalue** a has **geometric multiplicity** one the matrix $A - aI_3$ has rank two. In particular, $col(A - aI_3) \neq \mathbb{R}^3$. Let $\boldsymbol{w}$ be any vector in $\mathbb{R}^3$ which is not in $col(A - aI_3)$ and set $\boldsymbol{x} = (A - aI_3)\boldsymbol{w}, \boldsymbol{y} = (A - aI_3)\boldsymbol{x}$. It is our goal to show that $\boldsymbol{y}$ is an **eigenvector** for A and that $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is a **basis** for $\mathbb{R}^3$. Suppose we have shown this. Then we have the following:

$$A\boldsymbol{w} = a\boldsymbol{w} + \boldsymbol{x} \tag{7.25}$$

$$A\boldsymbol{x} = a\boldsymbol{x} + \boldsymbol{y} \tag{7.26}$$

$$A\boldsymbol{y} = a\boldsymbol{y} \tag{7.27}$$

The relation (7.27) follows from the definition of **eigenvalue** and the others from how we have defined $\boldsymbol{x}$ and $\boldsymbol{y}$.

We now establish the claim that $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is **linearly independent**. Suppose c_1, c_2, c_3 are scalars and $c_1\boldsymbol{w} + c_2\boldsymbol{x} + c_3\boldsymbol{y} = \boldsymbol{0}_3$. We need to show that $c_1 = c_2 = c_3 = 0$.

Multiply $c_1\boldsymbol{w} + c_2\boldsymbol{x} + c_3\boldsymbol{y}$ by A :

$$\boldsymbol{0}_3 = A\boldsymbol{0}_3 = A(c_1\boldsymbol{w} + c_2\boldsymbol{x} + c_3\boldsymbol{y}) + c_1A\boldsymbol{w} + c_2A\boldsymbol{x} + c_3A\boldsymbol{y} =$$

$$c_1(a\boldsymbol{w} + \boldsymbol{x}) + c_2(a\boldsymbol{x} + \boldsymbol{y}) + a\boldsymbol{y} = a(c_1\boldsymbol{w} + c_2\boldsymbol{x} + c_3\boldsymbol{y}) = (c_1\boldsymbol{x} + c_2\boldsymbol{y}) \tag{7.28}$$

Subtracting $a(c_1\boldsymbol{w} + c_2\boldsymbol{x} + c_3\boldsymbol{y}) = a\boldsymbol{0}_3 = \boldsymbol{0}_3$ from the last expression in (7.28) we obtain

$$c_1\boldsymbol{x} + c_2\boldsymbol{y} = \boldsymbol{0}_3 \tag{7.29}$$

Multiplying $c_1\boldsymbol{x} + c_2\boldsymbol{y} = \boldsymbol{0}_3$ by A we get

$$\boldsymbol{0}_3 = A\boldsymbol{0}_3 = A(c_1\boldsymbol{x} + c_2\boldsymbol{y}) = c_1A\boldsymbol{x} + c_2A\boldsymbol{y} =$$

$$c_1(a\boldsymbol{x} + \boldsymbol{y}) + c_2(a\boldsymbol{y}) = a(c_1\boldsymbol{x} + c_2\boldsymbol{y}) + c_1\boldsymbol{y} \tag{7.30}$$

We conclude from (7.30) that $c_1\boldsymbol{y} = 0$ and therefore $c_1 = 0$.

However, since $c_1\boldsymbol{x} + c_2\boldsymbol{y} = \boldsymbol{0}_3$ we conclude that $c_2\boldsymbol{y} = \boldsymbol{0}_3$ and hence $c_2 = 0$. Finally, since $c_1\boldsymbol{w} + c_2\boldsymbol{x} + c_3\boldsymbol{y} = \boldsymbol{0}_3$, it must be the case that $c_3\boldsymbol{y} = \boldsymbol{0}_3$ and therefore also $c_3 = 0$. Thus, $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is **linearly independent**, and, consequently, a **basis** for $\mathbb{R}^3$.

Since $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is a **basis** for $\mathbb{R}^3$, the matrix $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$ is **invertible** and

$$P^{-1}AP = \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 1 & a \end{pmatrix}.$$

So to complete this case we must show for any vector $\boldsymbol{w} \notin col(A - aI_3)$ that $(A - aI_3)^2\boldsymbol{w}$ is an **eigenvector** for A.

Since A and $A - aI_3$ commute, by **Lemma** (7.3.4) the column space of $A - aI_3$ is $A-$invariant. Since a is an **eigenvalue** with **geometric multiplicity** one, the rank of $A - aI_3$ is two. Therefore there is a non-zero vector $\boldsymbol{x} \in col(A - aI_3)$ which is not an **eigenvector** for A. Since $\boldsymbol{x}$ is not an **eigenvector**, $(\boldsymbol{x}, A\boldsymbol{x})$ is **linearly independent**. Now the sequence of three vectors, $(\boldsymbol{x}, A\boldsymbol{x}, A^2\boldsymbol{x})$ lies in $col(A - aI_3)$, which has **dimension two** and so is **linearly dependent**. Therefore there are scalars c_0, c_1, c_2 not all zero such that $c_2A^2\boldsymbol{x} + c_1A\boldsymbol{x} + c_0\boldsymbol{x} = (c_2A^2 + c_1A + c_0I_3)\boldsymbol{x} = \boldsymbol{0}_3$. Since $\boldsymbol{x}, A\boldsymbol{x}$ are linearly independent, $c_2 \neq 0$. Therefore, if we set $d_1 = \frac{c_1}{c_2}, d_0 = \frac{c_0}{c_2}$ then $(A^2 + d_1A + d_0I_3)\boldsymbol{x} = \boldsymbol{0}_3$.

We claim that the **eigenspace** E_a is contained in $col(A - aI_3)$. Suppose to the contrary that this is not the case. Then let $\boldsymbol{z}$ be an **eigenvector** for A and set $\boldsymbol{y} = A\boldsymbol{x}$. Then by **Theorem** (2.4.11) the sequence $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is **linearly independent** since $(\boldsymbol{x}, \boldsymbol{y})$ is **linearly independent** and $\boldsymbol{z} \notin Span(\boldsymbol{x}, \boldsymbol{y})$. Set $P = (\boldsymbol{z}\ \boldsymbol{x}\ \boldsymbol{y})$, an **invertible matrix**.

Since $A^2\boldsymbol{x} + d_1A\boldsymbol{x} + d_0\boldsymbol{x} = \boldsymbol{0}$, it follows that

$$A\boldsymbol{y} = -d_0\boldsymbol{x} - d_1\boldsymbol{y} \tag{7.31}$$

Set $B = P^{-1}AP$. Then from (7.30) and the fact that $\boldsymbol{z}$ is an **eigenvector** for A with **eigenvalue** a it follows that

$$B = \begin{pmatrix} a & 0 & 0 \\ 0 & 0 & -d_0 \\ 0 & 1 & -d_1 \end{pmatrix} \tag{7.32}$$

Since A and B are similar they have the same **characteristic polynomial**. The **characteristic polynomial** of B is $-(\lambda - a)(\lambda^2 + d_1\lambda + d_0)$ and therefore $\lambda^2 + d_1\lambda + d_0 = (\lambda - a)^2$. However, this implies that $(A - aI_3)^2\boldsymbol{x} = \boldsymbol{0}_3$ and so the vector $\boldsymbol{y}' = (A - aI_3)\boldsymbol{x}$ is an **eigenvector** for A contradicting the assumption that the **eigenspace** E_a is not contained in $col(A - aI_3)$.

Therefore it must be the case that the **eigenspace** E_a is contained in $col(A - aI_3)$. We can now finish the result.

Let $\boldsymbol{u}$ be an **eigenvector** for A and let $\boldsymbol{x}$ be a vector in $col(A - aI_3)$ which is not an **eigenvector** so $(\boldsymbol{x}, \boldsymbol{u})$ is **linearly independent** and therefore a **basis** for $col(A - aI_3)$. Since $col(A - aI_3)$ is $A-$ invariant, $A\boldsymbol{x}$ is a **linear combination** of $\boldsymbol{x}$ and $\boldsymbol{u}$ and so there are scalars b and c such that $A\boldsymbol{x} = b\boldsymbol{x} + c\boldsymbol{u}$. Note that $c \neq 0$ for otherwise $\boldsymbol{x}$ is an **eigenvector** which contradicts the assumption that A has a unique **eigenvalue** a and the **geometric multiplicity** of a is one.

Set $\boldsymbol{y} = c\boldsymbol{u}$ which is also an **eigenvector** . Then $A\boldsymbol{x} = b\boldsymbol{x} + \boldsymbol{y}$. Note that for any vector $d\boldsymbol{x} + e\boldsymbol{y}$ in $col(A - aI_3), (A - aI_3)(d\boldsymbol{x} + e\boldsymbol{y}) \neq \boldsymbol{x}$. Since $\boldsymbol{x} \in col(A - aI_3)$ there is a vector $\boldsymbol{w}$ such that $(A - aI_3)\boldsymbol{w} = \boldsymbol{x}$. It follows from the preceding remark that $\boldsymbol{w} \notin col(A - aI_3)$. It follows from **Theorem** (2.4.11) that $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is **linearly independent**. Consquently, $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is a **basis** for $\mathbb{R}^3$ and the matrix $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$ is **invertible**. Since

$$A\boldsymbol{w} = a\boldsymbol{w} + \boldsymbol{x}, A\boldsymbol{x} = b\boldsymbol{x} + \boldsymbol{y}, A\boldsymbol{y} = a\boldsymbol{y} \tag{7.33}$$

we can conclude that $B = P^{-1}AP = \begin{pmatrix} a & 0 & 0 \\ 1 & b & 0 \\ 0 & 1 & a \end{pmatrix}$. Since A and B are similar, $-(\lambda - a)^3 = \chi_A(\lambda) = \chi_B(\lambda) = -(\lambda - a)^2(\lambda - b)$ and therefore $b = a$ and this completes the theorem. $\square$

Definition 7.12. The matrices in (7.21) - (7.24) are referred to as the ***canonical forms for real*** 3×3 ***matrices***

What You Can Now Do

1. Find the **canonical form** B for a real 2×2 matrix A and determine a matrix P such that $P^{-1}AP = B$.

2. Find the **canonical form** B for a real 3×3 matrix A and determine a matrix P such that $P^{-1}AP = B$.

Method (How To Do It)

Method 7.3.1. Find the **canonical form** B for a real 2×2 matrix A and determine a matrix P such that $P^{-1}AP = B$.

First compute the **characteristic polynomial** $\chi_A(\lambda) = det(A - \lambda I_2)$.

Find the roots, using the quadratic formula if necessary, of $\chi_A(\lambda)$.

If the roots are a conjugate pair of complex numbers, $\alpha = a + bi, \overline{\alpha} = a - bi$ then the **canonical form** of A is $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$.

Now that the **canonical form** B of A has been determined we find an **invertible matrix** P such that $P^{-1}AP = B$. Begin by finding a **complex eigenvector** $\boldsymbol{z}$ with **eigenvalue** α :

Form the complex matrix $A - \alpha I_2$. Take the first row of this matrix and scale it so that the (1,1)-entry is one. If this row is $(1\ z)$ then $\boldsymbol{z} = \begin{pmatrix} -z \\ 1 \end{pmatrix}$ is an **eigenvector** for A with **eigenvalue** α.

Write $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ with $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$. Set $P = (\boldsymbol{x}\ \boldsymbol{y})$.

Suppose that the roots of $\chi_A(\lambda)$ are real. If the roots are a and b with $a \neq b$ then A is **diagonalizable** and has canonical form $B = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$.

To find P use **Method** (7.2.1).

Suppose that A has single real **eigenvalue** a with **algebraic multiplicity** two. If the **geometric multiplicity** is two then $A = aI_2$ and no matrix P is needed.

If the **geometric multiplicity** is one then A has **canonical form** $B = \begin{pmatrix} a & 0 \\ 1 & a \end{pmatrix}$. Choose a non-zero vector $\boldsymbol{w}$ which is not an **eigenvector** and set $\boldsymbol{v} = A\boldsymbol{w} - a\boldsymbol{w}$. Set $P = (\boldsymbol{w}\ \boldsymbol{v})$. This is the required matrix. To find $\boldsymbol{w}$ choose any nonzero vector $\boldsymbol{u}$ check to see if $\boldsymbol{u}$ is an **eigenvector** . If it is not, set $\boldsymbol{w} = \boldsymbol{u}$. If $\boldsymbol{u}$ is an **eigenvector** then choose any vector which is not an a multiple of $\boldsymbol{u}$ and set it equal to $\boldsymbol{w}$.

Example 7.3.4. Find the **canonical form** B of the matrix $A = \begin{pmatrix} 4 & 5 \\ -2 & -2 \end{pmatrix}$. Find an **invertible matrix** P such that $P^{-1}AP = B$.

The **characteristic polynomial** of A is $(4-\lambda)(-2-\lambda)+10=\lambda^2-2\lambda+2$. The roots of this polynomial are the conjugate pair of complex numbers $1+i, 1-i$. The **canonical form** of A is $B=\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. We find a matrix P such that $P^{-1}AP=B$.

$A-(1+i)I_2 = \begin{pmatrix} 3-i & 5 \\ -2 & -3-i \end{pmatrix}$ Scaling the first row by $(3-i)^{-1}$ it becomes $(1\ \frac{3}{2}+\frac{i}{2})$. The vector $\begin{pmatrix} -\frac{3}{2}-\frac{1}{2}i \\ 1 \end{pmatrix}$ is an eigenvector for A with eigenvalue $1+i$.

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

Example 7.3.5. Find the **canonical form** B of the matrix $A=\begin{pmatrix} 3 & 5 \\ -5 & -3 \end{pmatrix}$. Find an **invertible matrix** P such that $P^{-1}AP=B$.

The **characteristic polynomial** of A is $(3-\lambda)(-3-\lambda)+25=\lambda^2+16$. The roots of this polynomial are the conjugate pair $4i$ and $-4i$. The **canonical form** of A is the matrix $B=\begin{pmatrix} 0 & 4 \\ -4 & 0 \end{pmatrix}$. We find the matrix P.

$A-4iI_2 = \begin{pmatrix} 3-4i & 5 \\ -5 & -3-4i \end{pmatrix}$. Scaling the first row it becomes $(1\ \frac{3}{5}-\frac{4}{5}i)$. The vector $\boldsymbol{z}=\begin{pmatrix} -\frac{3}{5}-\frac{4}{5}i \\ 1 \end{pmatrix}$ is an **eigenvector** of A with **eigenvalue** $4i$.

$\begin{pmatrix} -\frac{3}{5}-\frac{4}{5}i \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{3}{5} \\ 1 \end{pmatrix} + \begin{pmatrix} -\frac{4}{5} \\ 0 \end{pmatrix} i$. The matrix $P=\begin{pmatrix} -\frac{3}{5} & -\frac{4}{5} \\ 1 & 0 \end{pmatrix}$.

Example 7.3.6. Find the **canonical form** B of the matrix $A=\begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$. Find an **invertible matrix** P such that $P^{-1}AP=B$.

The **characteristic polynomial** of A is $(3-\lambda)(1-\lambda)+1=\lambda^2-4\lambda+4=(\lambda-2)^2$. Therefore A has a single **eigenvalue** , two, with **algebraic multiplicity** two. Since A is not the scalar matrix $2I_2$ we know that A has **canonical form** $\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$. We find the matrix P.

We check to see if $\boldsymbol{e}_1=\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is an **eigenvector**. $A\boldsymbol{e}_1$ is the first column of A which is not a multiple of $\boldsymbol{e}_1$ and therefore $\boldsymbol{e}_1$ is not an **eigenvector**. Set $\boldsymbol{w}=\boldsymbol{e}_1$. Set

$\boldsymbol{v} = A\boldsymbol{e}_1 - 2\boldsymbol{e}_1 = \begin{pmatrix} 3 \\ -1 \end{pmatrix} - \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Check to see that this an **eigenvector** of A. Set $P = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$.

Method 7.3.2. Find the **canonical form** B for a real 3×3 matrix A and determine a matrix P such that $P^{-1}AP = B$.

Find the **characteristic polynomial** $\chi_A(\lambda)$ and find the roots of this polynomial. There are two major cases, some of which break down into multiple subcases. The two major cases are: I. The roots of $\chi_A(\lambda)$ are $a + bi, a - bi$ and c with c real. II. The **characteristic polynomial** has three real roots.

I. In this case the **canonical form** of A is the matrix (7.21). Use **Method** (7.1.3) to find an **eigenvector** $\boldsymbol{w}$ for the real **eigenvalue** c.

Next, find an **eigenvector** $\boldsymbol{z}$ for $\alpha = a + bi$. To do this, form the matrix $A - \alpha I_3$ and use **Method** (5.6.7) to find a complex vector which spans $null(A - \alpha I_3)$.

Now, write $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ and set $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$.

II. Use **Method** (7.2.1) to determine if A is **diagonalizable** and, if so, find a **diagonalizing matrix** P. In this case, A is in the **similarity class** of (7.22) and the **diagonalizing matrix** we obtain from that method is the matrix we seek.

Assume after applying **Method** (7.2.1) that A is not **diagonalizable matrix**. There are two possibilities: a) A has two distinct **eigenvalues** a with **algebraic multiplicity** two and **geometric multiplicity** one and $b \neq a$ with **algebraic multiplicity** one; and b) A has one **eigenvalue** a with **algebraic multiplicity** three and **geometric multiplicity** less than three.

a) In this case the **canonical form** of A is the matrix $B = \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 0 & b \end{pmatrix}$.

To find P, first use **Method** (7.1.3) to find an **eigenvector** $\boldsymbol{y}$ for the **eigenvalue** b.

Next, choose the first of the three columns of the matrix $C = A - bI_3$ which is not an **eigenvector** (for the **eigenvalue** a) and call this vector $\boldsymbol{w}$. Set $\boldsymbol{x} = A\boldsymbol{w} - a\boldsymbol{w}$. Then the matrix $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$ satisfies $P^{-1}AP = B$.

b) There are now two subcases: i) The **geometric multiplicity** of a is two; ii) the **geometric multiplicity** of a is one.

i) Suppose a has **geometric multiplicity** two. In this case A has **canonical form** $B = \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 0 & a \end{pmatrix}$.

To find P : Choose any vector $\boldsymbol{w}$ which is not in the two dimensional **eigenspace** E_a (at least one of the **standard basis vectors** will work). Set $\boldsymbol{x} = A\boldsymbol{w} - a\boldsymbol{w}$. Then use **Method** (7.1.3) to find a **basis** for E_a. At least one of these **basis vectors** for E_a is **not** a multiple of $\boldsymbol{x}$. Choose such a vector and label it $\boldsymbol{y}$. Set $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$.

ii) In this case A has **canonical form** $B = \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 1 & a \end{pmatrix}$.

To determine P : Find a vector $\boldsymbol{w}$ which is not in $null([A - aI_3]^2)$ (since $(A - aI_3)^2 \neq 0_{3\times 3}$ at least one of the **standard basis vectors** will work). Set $\boldsymbol{x} = A\boldsymbol{w} - a\boldsymbol{w}$ and $\boldsymbol{y} = A\boldsymbol{x} - a\boldsymbol{x}$. Then $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$.

Example 7.3.7. Find the **canonical form** B of the matrix $A = \begin{pmatrix} 1 & -2 & 0 \\ -2 & -3 & -4 \\ 0 & 4 & 3 \end{pmatrix}$ and find a matrix P such that $P^{-1}AP = B$.

The **characteristic polynomial** of A is

$$\chi_A(\lambda) = det \begin{pmatrix} 1-\lambda & -2 & 0 \\ -2 & -3-\lambda & -4 \\ 0 & 4 & 3-\lambda \end{pmatrix} =$$

$$(1-\lambda)(-3-\lambda)(3-\lambda) - (-2)(-2)(3-\lambda) - (4)(-4)(1-\lambda) =$$

$$-\lambda^3 + \lambda^2 + 9\lambda - 9 - 12 + 4\lambda + 16 - 16\lambda =$$

$$-\lambda^3 + \lambda^2 - 3\lambda - 5 = -(\lambda^3 - \lambda^2 + 3\lambda + 5) = -(\lambda + 1)(\lambda^2 - 2\lambda + 5).$$

The roots of $\chi_A(\lambda)$ are -1 and the complex conjugates $1 + 2i, 1 - 2i$.

We find an **eigenvector** for -1: $A - (-1)I_3 = A + I_3 = \begin{pmatrix} 2 & -2 & 0 \\ -2 & -2 & -4 \\ 0 & 4 & 4 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. We then deduce that the vector $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$ is an **eigenvector** of A with eigenvalue -1.

We next find a **complex eigenvector** for the **eigenvalue** $1 + 2i$.

$$A-(1+2i)I_3=\begin{pmatrix}-2i & -2 & 0\\ -2 & -4-2i & -4\\ 0 & 4 & 2-2i\end{pmatrix}$$

This matrix has **reduced echelon form** $\begin{pmatrix}1 & 0 & \frac{1}{2}+\frac{1}{2}i\\ 0 & 1 & \frac{1}{2}-\frac{1}{2}i\\ 0 & 0 & 0\end{pmatrix}$

Then $\boldsymbol{z}=\begin{pmatrix}-\frac{1}{2}-\frac{1}{2}i\\ -\frac{1}{2}+\frac{1}{2}i\\ 1\end{pmatrix}$ is an **eigenvector** for the **eigenvalue** $1+2i$. We can just as well multiply $\boldsymbol{z}$ by 2 to clear fractions so we can work with the vector $\boldsymbol{z}'=\begin{pmatrix}-1-i\\ -1+i\\ 2\end{pmatrix}$.

$\boldsymbol{z}'=\begin{pmatrix}-1\\ -1\\ 2\end{pmatrix}+\begin{pmatrix}-1\\ 1\\ 0\end{pmatrix}i=\boldsymbol{x}+\boldsymbol{y}i$. Set $P=\begin{pmatrix}-1 & -1 & -1\\ -1 & -1 & -1\\ 1 & 2 & 0\end{pmatrix}$. The $P^{-1}AP=$ $\begin{pmatrix}-1 & 0 & 0\\ 0 & 1 & 2\\ 0 & -2 & 1\end{pmatrix}$.

Example 7.3.8. Find the **canonical form** B of the matrix $A=\begin{pmatrix}1 & 3 & 2\\ 3 & 2 & 3\\ -2 & -3 & -3\end{pmatrix}$ and find a matrix P such that $P^{-1}AP=B$.

The **characteristic polynomial** of A is $\chi_A(\lambda)=$

$$det(A-\lambda I_3)=det\begin{pmatrix}1-\lambda & 3 & 2\\ 3 & 2-\lambda & 3\\ -2 & -3 & -3-\lambda\end{pmatrix}=$$

$$-\lambda^3+3\lambda+2=-(\lambda^3-3\lambda-2)=-(\lambda-2)(\lambda+1)^2.$$

So A has **eigenvalue** 2 with **algebraic multiplicity** one and -1 with **algebraic multiplicity** two. We find a **basis** for E_{-1} to determine if A is **diagonalizable matrix** .

$A-(-1)I_3=A+I_3=\begin{pmatrix}2 & 3 & 2\\ 3 & 3 & 3\\ -2 & -3 & -2\end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}$.
This matrix has **rank** two and **nullity** one and therefore we conclude that the **eigenvalue**

-1 has **geometric multiplicity** one and A is not **diagonalizable matrix**. Therefore the **canonical form** of A is $B = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. We find a matrix P such that $P^{-1}AP = B$.

Consider the matrix $C = A - 2I_3 = \begin{pmatrix} -1 & 3 & 2 \\ 3 & 0 & 3 \\ -2 & -3 & -5 \end{pmatrix}$. We find a column of C which is not an **eigenvector** for ..

$A \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} 4 \\ -3 \\ -1 \end{pmatrix}$ is not a multiple of $\begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}$. Set $\boldsymbol{w} = \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}$ and $\boldsymbol{x} = A\boldsymbol{w} - (-\boldsymbol{w}) = (A + I_3)\boldsymbol{w} = \begin{pmatrix} 3 \\ 0 \\ -3 \end{pmatrix}$, an **eigenvector** of A for the **eigenvalue** -1.

It remains to find an **eigenvector** with **eigenvalue** 2. We use **Method** (7.1.3):

$A - 2I_3 = \begin{pmatrix} -1 & 3 & 2 \\ 3 & 0 & 3 \\ -2 & -3 & -5 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$.

Therefore the vector $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$ is an **eigenvector** with **eigenvalue** 2. Now set $P = \begin{pmatrix} -1 & 3 & -1 \\ 3 & 0 & -1 \\ -2 & -3 & 1 \end{pmatrix}$. Then $P^{-1}AP = B$.

Example 7.3.9. Let $A = \begin{pmatrix} 3 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}$. Find the **canonical form** B of A and an **invertible matrix** P such that $P^{-1}AP = B$.

The **characteristic polynomial** of $A, \chi_A(\lambda) =$

$$det(A - \lambda I_3) = det \begin{pmatrix} 3-\lambda & 1 & -1 \\ 0 & 1-\lambda & 1 \\ 1 & 0 & 2-\lambda \end{pmatrix} =$$

$$-(\lambda^3 - 6\lambda^2 + 12\lambda - 8) = -(\lambda - 2)^3.$$

So, A has a single **eigenvalue** with **algebraic multiplicity** three. Since A is not $2I_3$, A is not **diagonalizable matrix**. We find the **geometric multiplicity** by **Method** (7.1.3) in order to find the **canonical form** and the matrix P.

$A - 2I_3 = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ and has **reduced echelon form** $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. This matrix has **rank** two and **nulllity** one so the **geometric multiplicity** of 2 is one. This implies that the **canonical form** is $B = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}$.

The **standard basis vector** $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is not an **eigenvector** so we can set $\boldsymbol{w} = \boldsymbol{e}_1$.

Set $\boldsymbol{x} = (A - 2I_3)\boldsymbol{w} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. Finally, set $\boldsymbol{y} = (A - 2I_3)\boldsymbol{x} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. Set $P = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$. Then $P^{-1}AP = B$.

Example 7.3.10. Let $A = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 0 & 1 \\ 1 & 0 & 4 \end{pmatrix}$. Find the **canonical form** B of A and an **invertible matrix** P such that $P^{-1}AP = B$.

The **characteristic polynomial** of $A, \chi_A(\lambda) = det(A - \lambda I_3) =$

$$(3 - \lambda)[(2 - \lambda)(4 - \lambda) + 1] = (3 - \lambda)[\lambda^2 - 6\lambda + 9] = -(\lambda - 3)^3.$$

So, 3 is an **eigenvalue** with **algebraic multiplicity** three. Since A is not the scalar matrix $3I_3$ it is not a **diagonalizable matrix**. We compute the **dimension** of $E_3 = null(A - 3I_3)$.

$A - 3I_3 = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. The **reduced echelon form** of this matrix is $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

This matrix has **rank** one and **nullity** two. $E_3 = Span\left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right)$.

The vector $\boldsymbol{e}_1$ is not in E_3 so set $\boldsymbol{w} = \boldsymbol{e}_1$. Set $\boldsymbol{x} = (A-3I_3)\boldsymbol{w} = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$. Finally, set $\boldsymbol{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Let $P = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. Then $P^{-1}AP = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.

Exercises

For each 2×2 real matrix A in exercises 1 - 12 find the **canonical form** B and an **invertible matrix** P such that $P^{-1}AP = B$. See **Method** (7.3.1).

1. $A = \begin{pmatrix} 2 & 5 \\ -1 & -2 \end{pmatrix}$.

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

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

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

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

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

7. $A = \begin{pmatrix} 4 & 5 \\ -1 & 2 \end{pmatrix}$

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

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

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

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

12. $A = \begin{pmatrix} 10 & 9 \\ -5 & -2 \end{pmatrix}$

In exercises 13 - 24 for each 3×3 real matrix A find the**canonical form** B and an **invertible matrix** P such that $P^{-1}AP = B$. See **Method** (7.3.2).

13. $A = \begin{pmatrix} 6 & -2 & -2 \\ 2 & -4 & 4 \\ 5 & -5 & 2 \end{pmatrix}$

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

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

16. $A = \begin{pmatrix} 2 & 3 & 5 \\ 0 & 1 & 1 \\ -1 & -2 & -3 \end{pmatrix}$

17. $A = \begin{pmatrix} -1 & -1 & 0 \\ 4 & -6 & 1 \\ 4 & -3 & -2 \end{pmatrix}$

18. $A = \begin{pmatrix} 0 & 0 & -4 \\ 3 & 2 & 6 \\ 1 & 0 & 4 \end{pmatrix}$

19. $A = \begin{pmatrix} -1 & -12 & 12 \\ 1 & 11 & -13 \\ 2 & 6 & -3 \end{pmatrix}$

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

21. $A = \begin{pmatrix} -12 & 15 & 7 \\ -6 & 8 & 3 \\ -10 & 12 & 7 \end{pmatrix}$

22. $A = \begin{pmatrix} 5 & 3 & 4 \\ -7 & -5 & -4 \\ -3 & -1 & -4 \end{pmatrix}$

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

24. $A = \begin{pmatrix} 7 & 4 & -2 \\ -2 & 1 & 1 \\ 4 & 4 & 1 \end{pmatrix}$

In exercises 25-29 answer true or false and give an explanation.

25. If A is a real $n \times n$ matrix with **eigenvalue** $3-4i$ and $-3+4i$ is also an eigenvalue of A.

26. If $\boldsymbol{x} = \begin{pmatrix} 1+i \\ 1-i \end{pmatrix}$ is an **eigenvector** of the real 2×2 matrix A with **eigenvalue** $2+3i$ then $\overline{\boldsymbol{x}} = \begin{pmatrix} 1-i \\ 1+i \end{pmatrix}$ is an **eigenvector** of A with **eigenvalue** $2-3i$.

27. If A is a complex **invertible matrix** then $\overline{A^{-1}} = \overline{A}^{-1}$.

28. If A is a real 3×3 matrix then A has a real **eigenvector**.

29. If A is a real 2×2 matrix then A is a **diagonalizable matrix** if and only if the **eigenvalues** of A are real numbers.

Challenge Exercises (Problems)

1. Assume that A is a 3×3 real matrix with **characteristic polynomial** $-(\lambda-a)^2(\lambda-b)$ where $a \neq b$ and the **eigenvalue** a has **geometric multiplicity** one.

a) Let $\boldsymbol{z}$ be an **eigenvector** of A with **eigenvalue** a. Prove the $\boldsymbol{z} \in col(A-bI_3)$. (Hint: Show that $(A-bI_3)\boldsymbol{z}$ is a multiple of $\boldsymbol{z}$.)

b) Explain why $col(A-bI_3)$ has **rank** two and is $A-$invariant.

c) Explain why you can choose non-zero vector $\boldsymbol{x} \in col(A-bI_3)$ which is **not** an **eigenvector** for A with **eigenvalue** a.

d) Set $\boldsymbol{y} = (A-aI_3)\boldsymbol{x}$. Prove that $\boldsymbol{y}$ is not a multiple of $\boldsymbol{x}$. Explain why there are scalars c and d such that $A\boldsymbol{y} = c\boldsymbol{x} + d\boldsymbol{y}$.

e) Let $\boldsymbol{w}$ be an **eigenvector** for A with **eigenvalue** b. Explain why $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is a **basis** for $\mathbb{R}^3$.

f) Set $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$. Explain why P is **invertible** and $P^{-1}AP = B = \begin{pmatrix} b & 0 & 0 \\ 0 & a & c \\ 0 & 1 & d \end{pmatrix}$.

g) Use the fact that the $\chi_B(\lambda) = \chi_A(\lambda) = -(\lambda-b)(\lambda-a)^2$ to conclude that $c = 0$ and $d = a$.

2. Let A be a real 3×3 matrix with **characteristic polynomial** $-(\lambda-a)^3$ and assume that the eigenvalue a has **geometric multiplicity** two.

a) Explain why $col(A - aI_3)$ has **rank** one.

b) Let $\boldsymbol{z}$ be any non-zero vector in $col(A - aI_3)$. Prove that $\boldsymbol{z}$ is an **eigenvector** for A (with **eigenvalue** a). Hint: $(A - aI_3)\boldsymbol{z}$ is in $col(A - aI_3)$ and therefore a multiple of $\boldsymbol{z}$. Prove if $(A - aI_3)\boldsymbol{z} \neq \mathbf{0}_3$ then A has an **eigenvalue** not equal to a.]

c) Let $\boldsymbol{w}$ be any vector not in E_a (the **eigenspace** for a. Set $\boldsymbol{x} = (A - aI_3)\boldsymbol{w}$. Explain why $\boldsymbol{x} \neq \mathbf{0}_3$ and is an **eigenvector** for A (with **eigenvalue** a).

d) Let $\boldsymbol{y}$ be any vector in E_a which is not a multiple of $\boldsymbol{x}$. Prove that $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y})$ is a **basis** for $\mathbb{R}^3$.

e) Set $P = (\boldsymbol{w}\ \boldsymbol{x}\ \boldsymbol{y})$. Explain why P is **invertible** and why $P^{-1}AP = \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 0 & a \end{pmatrix}$

Quiz Solutions

1. a) $\chi_A(\lambda) = \lambda^2 - 6\lambda + 13$. The roots are $3 + 2i, 3 - 2i$. Not right, see **Method** (7.1.1) and **the quadratic formula**.

1.b) $\begin{pmatrix} 2 & 5 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 1 - 2i \\ 1 \end{pmatrix} = \begin{pmatrix} 7 - 4i \\ 3 + 2i \end{pmatrix}$. Not right, see **definition of matrix multiplication**, **definition of multiplication of complex numbers**, and **definition of addition of complex numbers**.

1. c) $(3+2i) \begin{pmatrix} 1 - 2i \\ 1 \end{pmatrix} = \begin{pmatrix} 7 - 4i \\ 3 + 2i \end{pmatrix}$. Not right, see **definition of scalar multiplication of n-vectors**, **definition of multiplication of complex numbers**, and **definition of addition of complex numbers**.

2. $Span\left(\begin{pmatrix} -1 - i \\ 1 \end{pmatrix}\right)$. Not right, see **Method** (3.2.2).

3. $\chi_A(\lambda) = -\lambda^3 + 4\lambda^2 - 6\lambda + 4 = -(\lambda - 2)(\lambda^2 - 2\lambda + 2)$. The root are $\lambda = 2, \lambda = 1 + i$ and $\lambda = 1 - i$. Not right, see **Method** (7.1.1) and **the quadratic formula**.

4. The **reduced echelon form** of the martrix is $\begin{pmatrix} 1 & 0 & -1 + i \\ 0 & 1 & i \\ 0 & 0 & 0 \end{pmatrix}$. The **null space** is $Span\left(\begin{pmatrix} 1 - i \\ -i \\ 1 \end{pmatrix}\right)$. Not right, see **Method** (3.2.2).

7.4. How to Use It: Applications of Eigenvalues and Eigenvectors

In this section we make use of **eigenvalues** and **eigenvectors** in several applications. Some were already introduced such as **Markov chains** and **age structured population models** and we make use of **eigenvectors** and **eigenvalues** to achieve a deeper understanding. We also introduce some new applications, including linear recurrence relations, and linear differential systems.

Markov Chains

We begin with an investigation of **Markov chains** and apply the theory of **eigenvalues** and **eigenvectors** to prove some of the assertions we made earlier.

Recall that we made a number claims about **Markov chains** in Section (3.7). These were:

1. If P is an $n \times n$ **stochastic matrix** then P has a **steady state vector**. That is, there exists a **probability vector** $\boldsymbol{q}$ such that $P\boldsymbol{q} = \boldsymbol{q}$.

2. If every entry is positive for some power of P, P^k (we say P is **regular**) then P has a unique **steady state vector**.

3. If P is **regular** and $\boldsymbol{q}$ is the **steady state vector** for P then the powers P^k of P approach the matrix $S = (\boldsymbol{q}\ \boldsymbol{q}\ \ldots\ \boldsymbol{q})$.

In this subsection we will also show the following:

4. If $\{\boldsymbol{x}_k : k = 0, 1, \ldots\}$ is any **Markov chain** with **regular transition matrix** P and **steady state vector** $\boldsymbol{q}$ then $\boldsymbol{x}_k$ approaches $\boldsymbol{q}$.

Note, we have changed out notation from Section (3.7) and are using P to represent a **transition matrix**, this is because we will make use of the **transpose** repeatedly in this section and we find the notation T^{Tr} awkward.

In Section (3.7) we observed for each of our examples that P had a **steady state vector**. Such a vector we can now see is an **eigenvector** with **eigenvalue** 1 for the **transition matrix** P. We therefore first show that a **stochastic matrix** has **eigenvalue** 1.

Theorem 7.4.1. *Let P be an $n \times n$ **stochastic matrix**. Then 1 is an **eigenvalue** of P.*

Proof. We first prove that P^{Tr} has **eigenvalue** 1. Let $\boldsymbol{j}$ denote the row vector of length n with all entries 1. For any **probability vector**, $\boldsymbol{x}$, $\boldsymbol{j}\boldsymbol{x} = \boldsymbol{x} \cdot \boldsymbol{j}^{Tr} = 1$ since this is the sum of the entries of $\boldsymbol{x}$ which add to 1 by the definition of a probability vector.

Since the columns of a **stochastic matrix** P are all **probability vectors** it follows that $\boldsymbol{j}P = \boldsymbol{j}$ and consequently, $\boldsymbol{j}^{Tr} = (\boldsymbol{j}P)^{Tr} = P^{Tr}\boldsymbol{j}^{Tr}$ and therefore $\boldsymbol{j}^{Tr}$ is an **eigenvector** of P^{Tr} with eigenval 1. Since P and P^{Tr} have the same **characteristic polynomial**, P and P^{Tr} have the same **eigenvalues** (with the same **algebraic multiplicities**). Thus, 1 is an **eigenvalue** for P. $\square$

Let P be a **stochastic matrix** and let $\boldsymbol{x}$ be an **eigenvector** for P with **eigenvalue** 1. By multiplying $\boldsymbol{x}$ be a scalar, if necessary, we can assume that the sum of the entries in

$\boldsymbol{x}$ is 1. It is tempting to conclude that $\boldsymbol{x}$ is a **steady state vector** for P. However, we apply this terminology when $\boldsymbol{x}$ is a **probability vector**, which requires not only that the sum of its entries is 1, but also that the entries are all non-negative. At this point it is not obvious that there is such a vector and consequently some deeper analysis is needed. The crux of the argument will come from the following theorem which can be found in more advanced linear algebra texts as well as books on **Markov chains**:

Theorem 7.4.2. *Let P be an $n \times n$ **stochastic matrix** with **eigenvalue** λ. Then the following hold:*

1. $|\lambda| \leq 1$.

*2. If P is **regular** and $\lambda \neq 1$ then $|\lambda| < 1$.*

*3. If P is **regular** and **diagonalizable** then the **geometric multiplicity** of 1 as an **eigenvalue** for P^{Tr} is one.*

Example 7.4.1. Let $P = \begin{pmatrix} .8 & .4 \\ .2 & .6 \end{pmatrix}$. The **characteristic polynomial** of P is $\lambda^2 - 1.4\lambda + .4$ which factors as $(\lambda - 1)(\lambda - .4)$. The corresponding **eigenvectors** are $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

If we set $Q = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}$ then $Q^{-1}PQ = \begin{pmatrix} 1 & 0 \\ 0 & .4 \end{pmatrix}$ and therefore $P = Q\begin{pmatrix} 1 & 0 \\ 0 & .4 \end{pmatrix}Q^{-1}$. It then follows that

$$P^k = \begin{pmatrix} .8 & .4 \\ .2 & .6 \end{pmatrix}^k = Q\begin{pmatrix} 1 & 0 \\ 0 & .4 \end{pmatrix}^k Q^{-1} = Q\begin{pmatrix} 1 & 0 \\ 0 & (.4)^k \end{pmatrix}Q^{-1}.$$

As $k \to \infty$, $\begin{pmatrix} 1 & 0 \\ 0 & (.4)^k \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and therefore

$$P^k \to Q\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}Q^{-1} = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Observe that the columns of the limit matrix of P^k are identical; moreover this common vector is a **probability** vector and an **eigenvector** for P.

Additionally, for any **probability vector** $\boldsymbol{x}_0 = \begin{pmatrix} a \\ b \end{pmatrix}$, $P^k\boldsymbol{x}_0 \to \begin{pmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \frac{2}{3}a + \frac{2}{3}b \\ \frac{1}{3}a + \frac{1}{3}b \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$ since $a + b = 1$.

Under reasonable hypotheses, the phenomena which we observe in Example (7.4.1) holds in general. In particular, when P is **regular** and **diagonalizable**, P^k will approach a matrix $S = (\boldsymbol{q}\boldsymbol{q}\ldots\boldsymbol{q})$ where $\boldsymbol{q}$ is a **probability vector** and an **eigenvector** for P with **eigenvalue** 1.

Theorem 7.4.3. *Let P be a* **regular** *and* **diagonalizable** **stochastic matrix**. *Then the following hold:*

1. The **algebraic multiplicity** *and the* **geometric multiplicity** *of 1 as an* **eigenvalue** *for P is one.*

2. There is a **probability vector** $\boldsymbol{q}$ *such that $P^k \to (\boldsymbol{q}\ \boldsymbol{q}\ \ldots\ \boldsymbol{q})$ as $k \to \infty$. The vector $\boldsymbol{q}$ is an* **eigenvector** *for P with* **eigenvalue** *1.*

3. For any **probability vector** $\boldsymbol{x}_0$ *the vectors $\boldsymbol{x}_k = P^k\boldsymbol{x}_0$ approach $\boldsymbol{q}$ as $k \to \infty$.*

Proof. 1. Because P is **regular** we can apply **Theorem** (7.4.2). By the third part we can conclude that the **eigenvalue** 1 has **geometric multiplicity** one for P^{Tr}. Since P is **diagonalizable matrix**, so is P^{Tr}. It is then the case that the **algebraic multiplicity** of 1 as an **eigenvalue** for P^{Tr} is one, and therefore the **algebraic multiplicity** and the **geometric multiplicity** of 1 as an **eigenvalue** for P is one as well.

2. Since P is a **diagonalizable matrix**, there is a **basis** of $\mathbb{R}^n$ consisting of **eigenvectors** for P. Let $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a **basis** of **eigenvectors** with corresponding **eigenvalues** $(\lambda_1, \ldots, \lambda_n)$, where the notation is chosen so that $\lambda_1 = 1$. Set $Q = (\boldsymbol{x}_1\ \boldsymbol{x}_2 \ldots \boldsymbol{x}_n)$. Then $Q^{-1}PQ = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\} = D$.

Multiplying on the left by Q and the right by Q^{-1} we obtain

$$P = QDQ^{-1}$$

Taking each side to the k^{th} power we get

$$P^k = QD^kQ^{-1}$$

since for any $n \times n$ matrices A and B, $(QAQ^{-1})(QBQ^{-1}) = Q(AB)Q^{-1}$.

Now $D^k = diag\{1^k, \lambda_2^k, \dots, \lambda_n^k\}$. For $i \geq 2$ we have seen that $|\lambda_i| < 1$ and therefore as $k \to \infty$, $\lambda_i^k \to 0$. Therefore as $k \to \infty$, $D^k \to diag\{1, 0, \dots, 0\}$.

It follows that P^k approaches $Q \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} Q^{-1}$ as $k \to \infty$. Denote this limit matrix by S. Now for each k, P^k is a **stochastic matrix** and therefore S is a stochastic matrix. However, since the matrix $diag\{1, 0, \dots, 0\}$ has **rank** one, the limit of P^k has **rank** at most one. Since it is not the zero matrix (it is **stochastic**) it has **rank** exactly one. This means that the columns of this matrix are all multiples of the first column. However, the columns are all **probability vectors** and therefore they must all be equal.

Denote this vector by $\boldsymbol{q} = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}$.

Consider now the matrix $PS = (P\boldsymbol{q}\ P\boldsymbol{q}\ \dots\ P\boldsymbol{q})$. The limit of the sequence $P^{k+1} = P(P^k)$ of matrices will be PS. On the other hand, it is S. Therefore $S = PS$ and consequently $P\boldsymbol{q} = \boldsymbol{q}$. Thus, $\boldsymbol{q}$ is an **eigenvector** of P with **eigenvalue** 1. This completes 2.

3. Now let $\boldsymbol{x}_0$ be any **probability vector**. Then $P^k\boldsymbol{x}_0$ will approach $S\boldsymbol{x}_0 = (\boldsymbol{q}\,\boldsymbol{q}\ \dots\ \boldsymbol{q})\boldsymbol{x}_0$.

Assume that $\boldsymbol{x}_0 = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$. Then $S\boldsymbol{x}_0 = a_1\boldsymbol{q} + a_2\boldsymbol{q} + \dots a_n\boldsymbol{q} = (a_1 + a_2 + \dots + a_n)\boldsymbol{q} = 1\boldsymbol{q} = \boldsymbol{q}$. We have used the fact here that $\boldsymbol{x}_0$ is a **probability vector** and therefore that the sum of its entries is one. $\square$

Remark 7.10. When P is **regular** it is not necessary to assume that P is **diagonalizable** in order to obtain the results of **Theorem** (7.4.3) though the proof is considerably more complicated.

Age Structured Population Models

We return to the investigation of the **population model** introduced in Section (3.7). Recall that in this model we assumed that the species population is divided into n age groups, labeled by the natural numbers $i = 0, 1, 2, \dots, n-1$ where age group i consists of the individuals whose age a satisfies $i \leq a < i+1$. Typically, such models focus on the female population since they are the only ones which breed. The population of each group is to be followed in discrete units of time, typically the interval between breeding, usually quoted as years. The numbers of each age group are then recorded just prior to breeding. Moreover, the models incorporate the following assumptions:

1) There is a maximum age span for the species and consequently, every individual dies off by age n.

2) For each age group $i, i = 0, 1, 2, \dots, n-1$, there is a constant birth rate (fecundity), b_i, stated as average number of female off-spring per individual in the age group.

3) For each age group i there is a constant mortality rate, or the equivalent, a constant survival rate, s_i, which is the likelihood that an individual in the i^{th} group in period k survives to become a member of age group $i+1$ in period $k+1$.

All this data is encoded in the **population projection matrix**

$$S = \begin{pmatrix} b_1 & b_2 & \dots & b_{n-1} & b_n \\ s_1 & 0 & \dots & 0 & 0 \\ 0 & s_2 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & s_{n-1} & 0 \end{pmatrix}$$

We noted that this matrix had a particular shape and was an example of a type of matrix called a **Leslie matrix**. Because of the nature of this matrix we can draw some strong conclusions about the evolution of the species population over the long run. Before we state and prove our results we introduce some definitions and notation.

If $\boldsymbol{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}$ is a **population vector** (so all the entries are non-negative) then by $\overline{\boldsymbol{p}}$ we shall mean $p_1 + p_2 + \dots + p_n$ the ***total population***.

For a **population vector** $\boldsymbol{p}$ the ***associated age distribution vector*** is $\hat{\boldsymbol{p}} = \frac{1}{\overline{\boldsymbol{p}}}\boldsymbol{p}$. Notice that $\hat{\boldsymbol{p}}$ is a **probability vector**.

A **population vector** $\boldsymbol{p}$ will be said to have a ***stable age distribution*** if the age distribution vector of $\boldsymbol{p}$ and $S\boldsymbol{p}$ are identical. When this occurs, we have the following:

$$\frac{1}{\overline{S\boldsymbol{p}}}(S\boldsymbol{p}) = \frac{1}{\overline{\boldsymbol{p}}}\boldsymbol{p} \tag{7.34}$$

Multiplying both sides in (7.34) by $\overline{S\boldsymbol{p}}$ we obtain

$$S\boldsymbol{p} = \frac{\overline{S\boldsymbol{p}}}{\overline{\boldsymbol{p}}}\boldsymbol{p} \tag{7.35}$$

Thus, $S\boldsymbol{p}$ is a multiple of $\boldsymbol{p}$ and therefore an **eigenvector** of S. Note that the corresponding **eigenvalue** is $\frac{\overline{S\boldsymbol{p}}}{\overline{\boldsymbol{p}}}$ which is positive. When $\boldsymbol{p}$ has a stable age distribution the vector $\hat{\boldsymbol{p}}$ is called a ***stationary population distribution vector***. Because of the special nature of the **population projection matrix** S we can show that there is a unique stationary population distribution vector. Before doing so we first return to an example from Section (3.7).

Example 7.4.2. Consider the **population projection matrix**

$$S = \begin{pmatrix} 0 & 7 & 5 \\ .5 & 0 & 0 \\ 0 & .4 & 0 \end{pmatrix}$$

This matrix has **characteristic polynomial** $-(\lambda^3 - 3.5\lambda - 1) = -(\lambda - 2)(\lambda^2 + 2\lambda + .5)$ with roots 2, $-1 + \frac{\sqrt{2}}{2}$ and $-1 - \frac{\sqrt{2}}{2}$. The latter two roots are both negative and so 2 is the only positive **eigenvalue** . The vector $\boldsymbol{x}_1 = \begin{pmatrix} 20 \\ 5 \\ 1 \end{pmatrix}$ is an **eigenvector** for the **eigenvalue** 2. Notice that the entries of $\boldsymbol{x}_1$ are positive so that this can represent a **population vector** and this vector has a **stable age distribution**. Moreover, since the other two **eigenvalues** are negative, there are no other stationary population distribution vectors.

We now state and prove our result:

Theorem 7.4.4. *Let* $S = \begin{pmatrix} b_1 & b_2 & \dots & b_{n-1} & b_n \\ s_1 & 0 & \dots & 0 & 0 \\ 0 & s_2 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & s_{n-1} & 0 \end{pmatrix}$ *be a* **population projection matrix** *(Leslie matrix). Then S has a unique positive* **eigenvalue**. *Moreover this* **eigenvalue** *has a corresponding* **eigenvector** *with positive coordinates.*

Proof. The **characteristic polynomial** of S is

$$(-1)^n(\lambda^n - b_1\lambda^{n-1} - b_2s_1\lambda^{n-2} - b_3s_2s_1\lambda^{n-3} - \cdots - b_ns_{n-1}s_{n-2}\ldots s_1) =$$

$$(-1)^n f(\lambda).$$

We are assuming that all the survival rates are positive (otherwise the life span is less than n years). The birthrates are obviously nonnegative and we are assuming that at least some of the birthrates are positive. Therefore all the coefficients of the polynomial $f(\lambda) = \lambda^n - b_1\lambda^{n-1} - b_2s_1\lambda^{n-2} - b_3s_2s_1\lambda^{n-3} - \cdots - b_ns_{n-1}s_{n-2}\ldots s_1$, apart from λ^n are non-positive and at least one is negative. There is a **theorem on polynomials**, due to **Descartes**, which gives a bound on the number of positive roots by the number of sign changes among the coefficients. In the present case this is one. So, there is at most one positive root. We need to show that there is a positive root. We use the **intermediate value theorem** from calculus: If we can show there are positive values for λ for which the $f(\lambda)$ is positive and also positive values of λ for which $f(\lambda)$ is negative, then between such values there is a positive λ for which $f(\lambda)$ is zero.

For very large, positive λ, $f(\lambda) > 0$ and so we need to show there are positive λ such that $f(\lambda) < 0$. Let i be the first index such that $b_i \neq 0$. Then for positive λ, $f(\lambda) < \lambda^n - b_is_{i-1}s_{i-2}\ldots s_1\lambda^{n-i}$. Now $b_is_{i-1}s_{i-2}\ldots s_1 > 0$. Let $\alpha = \sqrt[i]{b_is_{i-1}s_{i-2}\ldots s_1}$ and choose λ so that $0 < \lambda < \alpha$. Then $f(\lambda) < \lambda^n - b_is_{i-1}s_{i-2}\ldots s_1\lambda^{n-i} < 0$ and therefore $f(\lambda)$ has a unique positive root. Denote this root by γ.

Consider the vector

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ \frac{s_1}{\gamma} \\ \frac{s_1s_2}{\gamma^2} \\ \frac{s_1s_2s_3}{\gamma^3} \\ \vdots \\ \frac{s_1s_2\ldots s_{n-1}}{\gamma^{n-1}} \end{pmatrix}.$$

This vector has positive coordinates. We claim that it is an **eigenvector** for S with **eigenvalue** γ. By straightforward multiplication we get

$$Sx = \begin{pmatrix} b_1 + b_2\frac{s_1}{\gamma} + b_3\frac{s_1 s_2}{\gamma^2} + \cdots + b_n\frac{s_1 s_2 \ldots s_{n-1}}{\gamma^{n-1}} \\ s_1 \\ \frac{s_1}{\gamma} s_2 \\ \frac{s_1 s_2}{\gamma^2} s_3 \\ \vdots \\ \frac{s_1 s_2 \ldots s_{n-2}}{\gamma^{n-2}} s_{n-1} \end{pmatrix} = \begin{pmatrix} b_1 + b_2\frac{s_1}{\gamma} + b_3\frac{s_1 s_2}{\gamma^2} + \cdots + b_n\frac{s_1 s_2 \ldots s_{n-1}}{\gamma^{n-1}} \\ \frac{s_1}{\gamma}\gamma \\ \frac{s_1 s_2}{\gamma^2}\gamma \\ \frac{s_1 s_2 s_3}{\gamma^3}\gamma \\ \vdots \\ \frac{s_1 s_2 \ldots s_{n-1}}{\gamma^{n-1}}\gamma \end{pmatrix} \tag{7.36}$$

Now γ is an **eigenvalue** and so it satisfies $f(\gamma) = 0$. From this we can conclude that

$$b_1\gamma^{n-1} + b_2 s_1 \gamma^{n-2} + b_3 s_2 s_1 \gamma^{n-3} + \cdots + b_n s_{n-1} s_{n-2} \ldots s_1 = \gamma^n \tag{7.37}$$

Dividing both sides by γ^{n-1} in (7.37) we obtain

$$b_1 + \frac{b_2 s_1}{\gamma} + \frac{b_3 s_2 s_1}{\gamma^2} + \cdots + \frac{b_n s_{n-1} s_{n-2} \ldots s_1}{\gamma^{n-1}} = \gamma \tag{7.38}$$

Substituting this into the last vector in (7.36) we obtain

$$Sx = \begin{pmatrix} \gamma \\ \frac{s_1}{\gamma}\gamma \\ \frac{s_1 s_2}{\gamma^2}\gamma \\ \frac{s_1 s_2 s_3}{\gamma^3}\gamma \\ \vdots \\ \frac{s_1 s_2 \ldots s_{n-1}}{\gamma^{n-1}}\gamma \end{pmatrix} = \gamma \begin{pmatrix} 1 \\ \frac{s_1}{\gamma} \\ \frac{s_1 s_2}{\gamma^2} \\ \vdots \\ \frac{s_1 s_2 \ldots s_{n-1}}{\gamma^{n-1}} \end{pmatrix} = \gamma x \tag{7.39}$$

Thus, x is an **eigenvector** with **eigenvalue** γ as claimed. □

Linear recurrence relations

You are probably familiar with the **Fibonacci sequence** $\{f_n\}_{n=1}^{\infty}$ which is the infinite sequence of natural numbers which begins with $f_0 = 0, f_1 = 1$ and where each subsequent term is expressed as the sum of the previous two:

$$f_n = f_{n-1} + f_{n-2}$$

for $n \geq 2$. This is an example of a recurrence relation of order 2. More general is the notion of a linear recurrence relation of order k.

Definition 7.13. Let k be a natural number. A sequence $\{s_n\}_{n=0}^{\infty}$ of real numbers is said satisfy a ***recurrence relation of order*** k if there are scalars $c_1, c_2, \ldots, c_k$ such that for every $n \geq k$

$$s_n = c_1 s_{n-1} + c_2 s_{n-2} + \ldots c_k s_{n-k} \tag{7.40}$$

It is straightforward to show using the **second principle of mathematical induction** (sometimes referred to as complete mathematical induction) that given scalars $c_1, c_2, \ldots, c_k$ and initial values $s_0, s_1, s_2, \ldots, s_{k-1}$ there is a unique sequence $\{s_n\}_{n=0}^{\infty}$ which satisfies the recurrence relation (7.40).

Suppose $\{s_n\}_{n=0}^{\infty}$ satisfies a recurrence relation of order one. Then there is scalar c such that $s_1 = cs_0, s_2 = cs_1 = c^2 s_0, s_3 = cs_2 = c(c^2 s_0) = c^3 s_0$. In general $s_n = c^n s_0$. We will see, making use of **diagonalization of matrices**, that many sequences which are given by **recurrence relations of order k** are sums of exponential functions. Before preceding to the general situation we illustrate with a couple of examples, one being the **Fibonacci sequence**.

Example 7.4.3. For the **Fibonacci sequence**, $\{f_n\}_{n=0}^{\infty}$, define the vectors $\boldsymbol{f}_n$ by

$$\boldsymbol{f}_n = \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix}$$

for $n = 0, 1, \ldots$. We define the recurrence matrix R for the **Fibonacci sequence** by

$$R = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

Let's see what we obtain when we multiply $\boldsymbol{f}_n$ by R :

$$R\boldsymbol{f}_n = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \begin{pmatrix} f_{n+1} \\ f_n + f_{n+1} \end{pmatrix} = \begin{pmatrix} f_{n+1} \\ f_{n+2} \end{pmatrix} = \boldsymbol{f}_{n+1} \tag{7.41}$$

The matrix R has **characteristic polynomial** $\lambda^2 - \lambda - 1$ with two distinct roots $\alpha_1 = \frac{1+\sqrt{5}}{2}$ and $\alpha_2 = \frac{1-\sqrt{5}}{2}$. Because the roots are simple (**algebraic multiplicity** one) R is **diagonalizable matrix** and there is a **basis** for $\mathbb{R}^2$ consisting of **eigenvectors** for R. The corresponding **eigenvectors** are $\boldsymbol{x}_1 = \begin{pmatrix} \frac{-1+\sqrt{5}}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} -\alpha_2 \\ 1 \end{pmatrix}$ for α_1 and $\boldsymbol{x}_2 = \begin{pmatrix} -\frac{1+\sqrt{5}}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} -\alpha_1 \\ 1 \end{pmatrix}$ for α_2.

Set $Q = (\boldsymbol{x}_1\ \boldsymbol{x}_2) = \begin{pmatrix} -\alpha_2 & -\alpha_1 \\ 1 & 1 \end{pmatrix}$. Then

$$Q^{-1}RQ = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}$$

from which we conclude that

$$R = Q \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} Q^{-1} \tag{7.42}$$

From (7.42) it follows that for every k that $R^k = [Q \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} Q^{-1}]^k = Q \begin{pmatrix} \alpha_1^k & 0 \\ 0 & \alpha_2^k \end{pmatrix} Q^{-1}$.

Since in general $\boldsymbol{f}_n = R^n \boldsymbol{f}_0 = R^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ it follows that

$$\begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \boldsymbol{f}_n = Q \begin{pmatrix} \alpha_1^n & 0 \\ 0 & \alpha_2^n \end{pmatrix} Q^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{7.43}$$

After performing the multiplications we get

$$f_n = \frac{1}{\sqrt{5}}(-\alpha_1^{n+1}\alpha_2 + \alpha_1\alpha_2^{n+1}) = \frac{1}{\sqrt{5}}(-\alpha_1\alpha_2)(\alpha_1^n - \alpha_2^n) = \frac{1}{\sqrt{5}}(\alpha_1^n - \alpha_2^n)$$

since $-\alpha_1\alpha_2 = 1$.

Example 7.4.4. Let $\{s_n\}_{n=0}^{\infty}$ be the sequence which satisfies $s_0 = 3, s_1 = 6, s_2 = 14$ and for $n \geq 3$

$$s_n = 6s_{n-1} - 11s_{n-2} + 6s_{n-3} \tag{7.44}$$

Set $\boldsymbol{f}_n = \begin{pmatrix} s_n \\ s_{n+1} \\ s_{n+2} \end{pmatrix}$ and $R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{pmatrix}$. Then

$$R\boldsymbol{f}_n = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{pmatrix} \begin{pmatrix} s_n \\ s_{n+1} \\ s_{n+2} \end{pmatrix} = \begin{pmatrix} s_{n+1} \\ s_{n+2} \\ 6s_{n+2} - 11s_{n+1} + 6s_n \end{pmatrix} = \begin{pmatrix} s_{n+1} \\ s_{n+2} \\ s_{n+3} \end{pmatrix} = \boldsymbol{f}_{n+1} \tag{7.45}$$

The last equality holds by the recurrence relation (7.44).

The **characteristic polynomial** of the matrix R is $(-1)(\lambda^3 - 6\lambda^2 + 11\lambda - 6)$ which has roots 1, 2 and 3. Since these are distinct the matrix R is **diagonalizable matrix**. The corresponding **eigenvectors** are:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 9 \end{pmatrix}$$

Let $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ and $Q = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix}$. Then

$$Q^{-1}RQ = D \tag{7.46}$$

Consequently, $R = QDQ^{-1}$ and $R^k = QD^kQ^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2^k & 0 \\ 0 & 0 & 3^k \end{pmatrix}$

In general $\boldsymbol{f}_n = R^n \boldsymbol{f}_0$ so by (7.46) we obtain

$$\boldsymbol{f}_n = Q \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2^n & 0 \\ 0 & 0 & 3^n \end{pmatrix} Q^{-1} \begin{pmatrix} 3 \\ 6 \\ 14 \end{pmatrix} \tag{7.47}$$

$Q^{-1} = \begin{pmatrix} 3 & -\frac{5}{2} & \frac{1}{2} \\ -3 & 4 & -1 \\ 1 & -\frac{3}{2} & \frac{1}{2} \end{pmatrix}$ and $Q^{-1} \begin{pmatrix} 3 \\ 6 \\ 14 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and therefore

$$\boldsymbol{f}_n = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2^n & 0 \\ 0 & 0 & 3^n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2^n \\ 3^n \end{pmatrix} = \begin{pmatrix} 1 + 2^n + 3^n \\ 1 + 2^{n+1} + 3^{n+1} \\ 1 + 2^{n+2} + 3^{n+2} \end{pmatrix}$$

Thus, in general $s_n = 1 + 2^n + 3^n$.

We now deal with the general situation. Before stating and proving our first theorem we make a definition suggested by **Example** (7.4.3) and **Example** (7.4.4).

Definition 7.14. Assume the infinite sequence $\{s_n\}_{n=0}^{\infty}$ satisfies the **k-order recurrence relation** $s_n = c_1 s_{n-1} + c_2 s_{n-2} + \dots c_k s_{n-k}$ for $n \geq k$. The ***recurrence matrix*** of the relation is the matrix $R =$

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ c_k & c_{k-1} & c_3 & \dots & c_2 & c_1 \end{pmatrix}.$$

We now prove our first theorem

Theorem 7.4.5. *Let $\{s_n\}_{n=0}^{\infty}$ satisfy the* **k-order recurrence relation** $s_n = c_1 s_{n-1} + c_2 s_{n-2} + \dots c_k s_{n-k}$ *for $n \geq k$. Assume that the polynomial $\lambda^k - c_1\lambda^{k-1} - c_2\lambda^{k-2} - \dots - c_k$ factors as $(\lambda - \alpha_1)(\lambda - \alpha_2)\dots(\lambda - \alpha_k)$ with the α_i distinct. Then there are scalars $b_1, b_2, \dots, b_k$ such that $s_n = b_1\alpha_1^n + b_2\alpha_2^n + \dots + b_k\alpha_k^k$.*

Proof. Define the vector $\boldsymbol{f}_n = \begin{pmatrix} s_n \\ s_{n+1} \\ \vdots \\ s_{n+k-1} \end{pmatrix}$ for $n \geq 0$ and let R be the **recurrence matrix** of the sequence, $R =$

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ c_k & c_{k-1} & c_3 & \dots & c_2 & c_1 \end{pmatrix}$$

Then

$$Rf_n = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ c_k & c_{k-1} & c_{k-2} & \dots & c_2 & c_1 \end{pmatrix} \begin{pmatrix} s_n \\ s_{n+1} \\ \vdots \\ s_{n+k-2} \\ s_{n+k-1} \end{pmatrix} =$$

$$\begin{pmatrix} s_{n+1} \\ s_{n+2} \\ \vdots \\ s_{n+k-1} \\ c_k s_n + c_{k-1} s_{n+1} + \dots c_1 s_{n+k-1} \end{pmatrix} = \begin{pmatrix} s_{n+1} \\ s_{n+2} \\ \vdots \\ s_{n+k-1} \\ s_{n+k} \end{pmatrix} = f_{n+1} \tag{7.48}$$

It then follows that $f_n = R^n f_0$.

The **characteristic polynomial** of the matrix R is $(-1)^k(\lambda^k - c_1\lambda^{k-1} - c_2\lambda^{k-2} - \dots - c_k)$ which by assumption has simple roots, $\alpha_1, \alpha_2, \dots, \alpha_k$. Therefore the matrix R is **diagonalizable matrix**. We remark that the each of the sequences $\{\alpha_i^n\}_{n=0}^{\infty}$ satisfy the recurrence. It follows from this that the vector $x_i = \begin{pmatrix} 1 \\ \alpha_i \\ \alpha_i^2 \\ \vdots \\ \alpha_i^{k-1} \end{pmatrix}$ is an **eigenvector** for R with eigenvalue α_i.

Set $Q = (x_1 \; x_2 \dots \; x_k)$ which is **invertible**. Then

$$Q^{-1}RQ = diag(\alpha_1, \alpha_2, \dots, \alpha_k) = D \tag{7.49}$$

It follows from (7.49) that $R = QDQ^{-1}$. Consequently,

$$R^n = QD^nQ^{-1} =$$

$$Qdiag(\alpha_1^n, \alpha_2^n, \dots, \alpha_k^n)Q^{-1} = Q \begin{pmatrix} \alpha_1^n & 0 & \dots & 0 \\ 0 & \alpha_2^n & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_k^n \end{pmatrix} Q^{-1} \tag{7.50}$$

Now $f_n = R^n f_0 = Qdiag(\alpha_1^n, \alpha_2^n, \dots, \alpha_k^n)Q^{-1} f_0$. Therefore each component of f_n is a **linear combination** of $\alpha_1^n, \alpha_2^n, \dots, \alpha_k^n$, in particular the first component, which is s_n. This completes the theorem. □

The situation when there are repeated roots is more complicated. We illustrate for a sequence of order 2.

Example 7.4.5. Let $\{s_n\}_{n=0}^{\infty}$ be a sequence which satisfies $s_0 = 3, s_1 = 8$ and for $n \geq 2$,

$$s_n = 4s_{n-1} - 4s_{n-2} \tag{7.51}$$

The **recursion matrix** for this sequence is $R = \begin{pmatrix} 0 & 1 \\ -4 & 4 \end{pmatrix}$ which has **characteristic polynomial** $\lambda^2 - 4\lambda + 4 = (\lambda - 2)^2$ and therefore it is not **diagonalizable matrix**.

The vector $\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is an **eigenvector** for the **eigenvalue** 2. Set $\boldsymbol{x}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $Q = (\boldsymbol{x}_1\ \boldsymbol{x}_2)$. Then $Q^{-1}RQ = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$. Set $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ so that $R = QAQ^{-1}$ and $R^n = QA^nQ^{-1}$.

It is an easy calculation to check that $A^n = \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix}$.

Then $\boldsymbol{f}_n = R^n \begin{pmatrix} 3 \\ 8 \end{pmatrix} = Q \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix} Q^{-1} \begin{pmatrix} 3 \\ 8 \end{pmatrix} =$

$$\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2^n & n2^{n-1} \\ 0 & 2^n \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 8 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2^n + n \cdot 2^n \\ 3 \cdot 2^{n+1} + (n+1)2^{n+1} \end{pmatrix}.$$

Therefore, $s_n = 3 \cdot 2^n + n2^n$.

The general case is summarized in the following theorem (which we will not prove).

Theorem 7.4.6. *Assume that the sequence $\{s_n\}_{n=0}^{\infty}$ satisfies the* **k-order recurrence relation**

$$s_n = c_1 s_{n-1} + c_2 s_{n-2} + \cdots + c_k s_{n-k}$$

and that the polyomial $\lambda^k - c_1\lambda^{k-1} - c_2\lambda^{k-2} - \cdots - c_{k-1}\lambda - c_k$ factors in the form $(\lambda - \alpha_1)^{e_1}(\lambda - \alpha_2)^{e_2} \ldots (\lambda - \alpha_t)^{e_t}$ where α_i are distinct and $e_1 + e_2 + \cdots + e_t = k$. Then $s_n = (b_{11}\alpha_1^n + b_{12}n\alpha_1^n + \cdots + b_{1e_1}n^{e_1-1}\alpha_1^n) + (b_{21}\alpha_2^n + b_{22}n\alpha_2^n + \cdots + b_{2e_2}n^{e_2-1}\alpha_2^n) + \cdots + (b_{t1}\alpha_t^n + b_{t2}n\alpha_t^n + \cdots + b_{te_t}n^{e_t-1}\alpha_t^n)$ for unique scalars b_{ij}.

Systems of linear differential equations

We begin with a definition.

Definition 7.15. Let $x_1, x_2, \ldots, x_n$ be differentiable functions of a variable t and denote by $x_i' = \frac{dx_i}{dt}$. By a ***system of first order linear differential equations*** we mean a system of equations in which each x_i' is expressed as a **linear combination** of $x_1, x_2, \ldots, x_n$:

$$\begin{array}{ccccccccc} x_1' & = & a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n \\ x_2' & = & a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & & \\ x_n' & = & a_{n1}x_1 & + & a_{n2}x_2 & + & \ldots & + & a_{nn}x_n \end{array} \tag{7.52}$$

Suppose we are given a system of first order linear differential equations (7.52).

Setting $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $\boldsymbol{x}' = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}$ and $A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$ then the system (7.52) can be written as

$$\boldsymbol{x}' = A\boldsymbol{x} \tag{7.53}$$

In some instances the system is quite simple to solve as the next example ilustrates.

Example 7.4.6. Solve the system of linear differential equations

$$x_1' = -2x_1$$

$$x_2' = 3x_2$$

$$x_3' = 4x_3$$

We know from calculus that the general solution to the differential equation $\frac{dx}{dt} = kx$ is Ce^{kt} for an arbitrary constant C. Therefore the solution to the given system is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} C_1 e^{-2t} \\ C_2 e^{3t} \\ C_3 e^{4t} \end{pmatrix}$$

As the next example will illustrate, this can be exploited when the matrix A of (7.53) is **diagonalizable matrix**.

Example 7.4.7. Solve the following **system of linear differential equations**

$$\boldsymbol{x}' = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} 4 & -1 & 1 \\ 12 & -9 & 7 \\ 12 & -12 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

The matrix $A = \begin{pmatrix} 4 & -1 & 1 \\ 12 & -9 & 7 \\ 12 & -12 & 10 \end{pmatrix}$ has **characteristic polynomial** $-(\lambda^3 - 5\lambda^2 - 2\lambda + 24) = -(\lambda+2)(\lambda-3)(\lambda-4)$. Therefore the **eigenvalues** of A are -2, 3, 4, which are all real and simple. Consequently, the matrix A is **diagonalizable matrix**. The corresponding **eigenvectors** are:

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

Let $Q = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix}$. Then $Q^{-1}AQ = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix} = D$ and therefore,

$$A = QDQ^{-1} \tag{7.54}$$

Now set $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = Q^{-1}\boldsymbol{x}$. Then $\boldsymbol{y}' = Q^{-1}\boldsymbol{x}' = Q^{-1}(A\boldsymbol{x})$.

By (7.54), $Q^{-1}(A\boldsymbol{x}) = Q^{-1}[(QDQ^{-1})\boldsymbol{x}] = (Q^{-1}Q)D(Q^{-1}\boldsymbol{x}) = D\boldsymbol{y} = \begin{pmatrix} -2y_1 \\ 3y_2 \\ 4y_3 \end{pmatrix}$

We have seen that the general solution to this system is $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} C_1 e^{-2t} \\ C_3 e^{3t} \\ C_3 e^{4t} \end{pmatrix}$.

Thus $Q^{-1}\boldsymbol{x} = \begin{pmatrix} C_1e^{-2t} \\ C_2e^{3t} \\ C_3e^{4t} \end{pmatrix}$ and consequently

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = Q\boldsymbol{y} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} C_1e^{-2t} \\ C_2e^{3t} \\ C_3e^{4t} \end{pmatrix} = \begin{pmatrix} C_2e^{3t} + C_3e^{4t} \\ C_1e^{-2t} + C_2e^{3t} + 2C_3e^{4t} \\ C_1e^{-2t} + 2C_3e^{4t} \end{pmatrix}$$

The techniques used in **Example** (7.4.7) can be applied in general when the matrix A of (7.53) is **diagonalizable**, which we next show.

Theorem 7.4.7. *Let A be an $n \times n$* **diagonalizable matrix** **similar** *to the matrix $D = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{pmatrix}$ via the matrix Q, that is. $Q^{-1}AQ = D$. Then the general solution to the* **system of linear differential equations**

$$\boldsymbol{x}' = A\boldsymbol{x}$$

is

$$\boldsymbol{x} = Q \begin{pmatrix} C_1e^{\alpha_1 t} \\ C_2e^{\alpha_2 t} \\ \vdots \\ C_ne^{\alpha_n t} \end{pmatrix}$$

Proof. Let $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = Q^{-1}\boldsymbol{x}$. Then $\boldsymbol{y}' = Q^{-1}\boldsymbol{x}'$. By hypothesis, $Q^{-1}AQ = D$ and so $A = QDQ^{-1}$. Therefore

$$\boldsymbol{y}' = Q^{-1}[(QDQ^{-1})\boldsymbol{x}] = (Q^{-1}Q)D[Q^{-1}\boldsymbol{x}] = D\boldsymbol{y} =$$

$$\begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{pmatrix} \boldsymbol{y} \tag{7.55}$$

The general solution to (7.55) is

$$\boldsymbol{y} = \begin{pmatrix} C_1 e^{\alpha_1 t} \\ C_2 e^{\alpha_2 t} \\ \vdots \\ C_n e^{\alpha_n t} \end{pmatrix} \tag{7.56}$$

From (7.56) we conclude that

$$\boldsymbol{x} = Q\boldsymbol{y} = Q \begin{pmatrix} C_1 e^{\alpha_1 t} \\ C_2 e^{\alpha_2 t} \\ \vdots \\ C_n e^{\alpha_n t} \end{pmatrix}$$

□

If $\boldsymbol{x}' = A\boldsymbol{x}$ is a **system of first order linear differential equations** by an ***initial condition*** we mean an assignment of $\boldsymbol{x}(0) = \boldsymbol{x}_0 = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$. When an iniitial condition is given we can determine the coefficients of the functions $e^{\alpha_i t}$ and get a unique solution for the system.

Example 7.4.8. Find the solution to the **system of first order linear differential equations**

$$\boldsymbol{x}' = \begin{pmatrix} 0 & -1 & -1 \\ -4 & -3 & -2 \\ 6 & 6 & 5 \end{pmatrix} \boldsymbol{x}$$

subject to the initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$

The **characteristic polynomial** of the matrix $A = \begin{pmatrix} 0 & -1 & -1 \\ -4 & -3 & -2 \\ 6 & 6 & 5 \end{pmatrix}$ is

$-(\lambda^3 - 2\lambda^2 - \lambda + 2) = -(\lambda - 1)(\lambda + 1)(\lambda - 2)$ and so the **eigenvalues** are 1, -1 and 2. The corresponding **eigenvectors** are $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$. Therefore, if we set $Q = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & -2 \end{pmatrix}$ then $Q^{-1}AQ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Then the general solution to the system $\boldsymbol{x}' = A\boldsymbol{x}$ is

$$\boldsymbol{x} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & -2 \end{pmatrix} \begin{pmatrix} C_1 e^t \\ C_2 e^{-t} \\ C_3 e^{2t} \end{pmatrix} = \begin{pmatrix} C_1 e^t + C_3 e^{2t} \\ -C_1 e^t + C_2 e^{-t} \\ -C_2 e^{-t} - 2C_3 e^{2t} \end{pmatrix} \tag{7.57}$$

To find the specific solution which satisfies the initial condition we set $\boldsymbol{x}(0) = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

From (7.57) we get

$$\boldsymbol{x}(0) = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & -2 \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} \tag{7.58}$$

It then follows that $\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = Q^{-1} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 2 & 2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}$.
Thus, the solution which satisfies the given initial condition is

$$\boldsymbol{x} = \begin{pmatrix} 2e^t - e^{2t} \\ -2e^t + 3e^{-t} \\ -3e^{-t} + 2e^{2t} \end{pmatrix}$$

Exercises

In exercises 1- 4 for the given **stochastic matrix** P : a) Determine the **eigenvalues** and corresponding **eigenvectors** ; b) Find a matrix Q such that $Q^{-1}PQ$ is a **diagonal matrix**; c) Find the long range transition matrix of P (the limit of P^k as $k \to \infty$).

1. $P = \begin{pmatrix} .9 & .3 \\ .1 & .7 \end{pmatrix}$

2. $P = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix}$

3. $P = \begin{pmatrix} .9 & .4 \\ .1 & .6 \end{pmatrix}$

4. $P = \begin{pmatrix} .6 & .2 & .2 \\ .2 & .6 & .2 \\ .2 & .2 & .6 \end{pmatrix}$

In exercises 5 - 9 determine the unique positive **eigenvalue** α of the given **population projection matrix** (Leslie matrix). Find a positive **eigenvector** for α and use this to compute a **stationary population distribution vector** $\hat{\boldsymbol{p}}$ for this matrix. Finally, determine if the population is increasing, decreasing or remains constant for this vector.

5. $\begin{pmatrix} 1 & 5 \\ .4 & 0 \end{pmatrix}$

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

7. $\begin{pmatrix} 0 & 4 & 12 \\ .2 & 0 & 0 \\ 0 & .16 & 0 \end{pmatrix}$

8. $\begin{pmatrix} 0 & 2 & 8 \\ .2 & 0 & 0 \\ 0 & .12 & 0 \end{pmatrix}$

In exercises 9 - 15 determine a closed formula for the given **linear recurrence**.

9. $\{s_n\}_{n=0}^{\infty}$ where $s_0 = 0, s_1 = 2$ and for $n \geq 2, s_n = 6s_{n-1} - 8s_{n-2}$.

10. $\{s_n\}_{n=0}^{\infty}$ where $s_0 = 0, s_1 = 2$ and for $n \geq 2, s_n = 4s_{n-1} - 3s_{n-2}$.

11. $\{s_n\}_{n=0}^{\infty}$ where $s_0 = 3, s_1 = 7, s_2 = 21$ and for $n \geq 3, s_n = 7s_{n-1} - 14s_{n-2} + 8s_{n-3}$

12. $\{s_n\}_{n=0}^{\infty}$ where $s_0 = 12, s_1 = 18$ and for $n \geq 2, s_n = 2s_{n-1} + 15s_{n-2}$.

13. $\{s_n\}_{n=0}^{\infty}$ where $s_0 = 0, s_1 = 1, s_2 = 6$ and for $n \geq 3, s_n = 3s_{n-1} - 3s_{n-2} + s_{n-3}$

14. $\{s_n\}_{n=0}^{\infty}$ where $s_0 = 5, s_1 = 16$ and for $n \geq 2, s_n = 4s_{n-1} - 4s_{n-2}$.

15. $\{s_n\}_{n=0}^{\infty}$ where $s_0 = 1, s_1 = 3, s_2 = 8$ and for $n \geq 3, s_n = 4s_{n-1} - 5s_{n-2} + 2s_{n-3}$.

In exercises 16 - 20 find the general solution to the **system of first order linear differential equations** $\boldsymbol{x}' = A\boldsymbol{x}$ for the given matrix A.

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

17. $A = \begin{pmatrix} 3 & 5 \\ -1 & -3 \end{pmatrix}$

18. $A = \begin{pmatrix} 5 & 4 \\ -2 & -1 \end{pmatrix}$

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

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

21. $A = \begin{pmatrix} 0 & -10 & -14 \\ 2 & -8 & -6 \\ -1 & 5 & 5 \end{pmatrix}$

In exercises 22 - 25 find the solution to the **system of first order linear differential equations** $\boldsymbol{x}' = A\boldsymbol{x}$ for the given matrix A and for the given **initial condition** $\boldsymbol{x}(0) = \boldsymbol{x}_0$.

22. $A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix}$, $\boldsymbol{x}_0 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$

23. $A = \begin{pmatrix} -4 & 3 \\ -2 & 3 \end{pmatrix}$, $\boldsymbol{x}_0 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

24. $A = \begin{pmatrix} 9 & 8 & -8 \\ -8 & 7 & -6 \\ 4 & -4 & 5 \end{pmatrix}$, $\boldsymbol{x}_0 = \begin{pmatrix} 6 \\ -7 \\ 4 \end{pmatrix}$

25. $A = \begin{pmatrix} -7 & -2 & 3 \\ 2 & -2 & -2 \\ -4 & -2 & 0 \end{pmatrix}$, $\boldsymbol{x}_0 = \begin{pmatrix} 4 \\ 0 \\ 5 \end{pmatrix}$

Chapter 8

Orthogonality in $\mathbb{R}^n$

8.1. Orthogonal and Orthonormal Sets in $\mathbb{R}^n$

In Section (2.6) we introduced the **dot product** of two vectors in $\mathbb{R}^n$ and a host of related and derivative concepts such as the **length (norm, magnitude)** of a vector, the **distance** and **angle** between two vectors and the notion that two vectors are **orthgonal**. In this chapter we study orthogonality in $\mathbb{R}^n$ in greater detail. In the current section we introduce several important concepts - orthogonal and orthonormal sets of vectors and orthonormal bases for **subspaces** of $\mathbb{R}^n$.

Am I Ready for This Material

The following are important to an understanding of this section

dot product

orthogonal or perpendicular vectors

length, norm or magnitude of a vector $\boldsymbol{v} \in \mathbb{R}^n$

distance between two vectors

unit vector

angle between two vectors

Quiz

1. In each of the following determine if the pair of vectors is **orthogonal**.

a) $\left(\begin{pmatrix} 1 \\ 2 \\ -3 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ -4 \\ -3 \\ -5 \end{pmatrix}\right)$

b) $\left(\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} -4 \\ -3 \\ 2 \\ 1 \end{pmatrix}\right)$

c) $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ -4 \\ -5 \\ 2 \end{pmatrix}\right)$

d) $\left(\begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -3 \\ 2 \\ 6 \\ -3 \end{pmatrix}\right)$

2. In each of the following vectors $\boldsymbol{v}$ find a **unit vector** with the same direction as $\boldsymbol{v}$.

a) $\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$

b) $\begin{pmatrix} 1 \\ -2 \\ 4 \\ -2 \end{pmatrix}$

c) $\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$

d) $\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$

3. a) Show that the sequence $\mathcal{S} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \right)$ is **linearly independent**.

b) Show that $\boldsymbol{u} = \begin{pmatrix} 6 \\ 4 \\ 2 \\ 0 \end{pmatrix}$ is in $Span(\mathcal{S})$ and determine the **coordinate vector** of $\begin{pmatrix} 6 \\ 4 \\ 2 \\ 0 \end{pmatrix}$ with respect to $\mathcal{S}$.

Quiz Solutions

New Concepts

In this section we introduce the following concepts

orthogonal sequence in $\mathbb{R}^n$

orthogonal basis for a subspace W of $\mathbb{R}^n$

orthonormal sequence in $\mathbb{R}^n$

orthornormal basis for a subspace W of $\mathbb{R}^n$

orthonormal matrix

orthogonal matrix

Theory (Why It Works)

We begin with an example:

Example 8.1.1. a) Show that the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are mutually **orthogonal** where

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

b) Prove that the sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a **basis** for $\mathbb{R}^3$.

c) Find the **coordinate vector** of $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ with respect to $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$.

a) $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = (1)(2) + (1)(-1) + (1)(-1) = 0;$

$\boldsymbol{v}_1 \cdot \boldsymbol{v}_3 = (1)(0) + (1)(1) + (1)(-1) = 0;$

$\boldsymbol{v}_2 \cdot \boldsymbol{v}_3 = (2)(0) + (-1)(1) + (-1)(-1) = 0.$

b) We could use **Gaussian elimination** to obtain an **echelon form** of the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3)$ to show that it is **invertible**, which would imply that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a basis of $\mathbb{R}^3$ by **Theorem** (3.4.15) but we give a non-computational argument.

Quite clearly, $\boldsymbol{v}_2$ is not a multiple of $\boldsymbol{v}_1$ and therefore $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is **linearly independent**. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly dependent**, then $\boldsymbol{v}_3$ must be a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ by part 2) of **Theorem** (2.4.7). So assume that $\boldsymbol{v}_3$ is a linear combination of $(\boldsymbol{v}_1, \boldsymbol{v}_2)$, say $\boldsymbol{v}_3 = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2$.

Then $\boldsymbol{v}_3 \cdot \boldsymbol{v}_3 = \boldsymbol{v}_3 \cdot (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1(\boldsymbol{v}_3 \cdot \boldsymbol{v}_1) + c_2(\boldsymbol{v}_3 \cdot \boldsymbol{v}_2)$ by **additivity and the scalar properties** of the **dot product**.

By a) $\boldsymbol{v}_3 \cdot \boldsymbol{v}_1 = \boldsymbol{v}_3 \cdot \boldsymbol{v}_2 = 0$ and therefore $\boldsymbol{v}_3 \cdot \boldsymbol{v}_3 = 0$. But then by **positive definiteness of the dot product**, $\boldsymbol{v}_3 = \boldsymbol{0}_3$, a contradiction. Therefore $\boldsymbol{v}_3$ is not a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ and $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is **linearly independent**. By **Theorem** (2.4.4) it follows that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a **basis** for $\mathbb{R}^3$.

c) We could find the **coordinate vector** of $\boldsymbol{u}$ by finding the **reduced echelon form** of the matrix $(\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3 \mid \boldsymbol{u})$ but we instead make use of the information we obtained in a).

Write $\boldsymbol{u} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3$ and take the **dot product** of $\boldsymbol{u}$ with $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$, respectively:

$$\boldsymbol{u}_1 \cdot \boldsymbol{v}_1 = (a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3) \cdot \boldsymbol{v}_1 = a_1(\boldsymbol{v}_1 \cdot \boldsymbol{v}_1) + a_2(\boldsymbol{v}_2 \cdot \boldsymbol{v}_1) + a_3(\boldsymbol{v}_3 \cdot \boldsymbol{v}_1) \quad (8.1)$$

by additivity and the scalar property of the dot product.

However, we showed in a) that $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are mutually **orthogonal**. Making use of this in Equation (8.1) we get

$$\boldsymbol{u} \cdot \boldsymbol{v}_1 = a_1(\boldsymbol{v}_1 \cdot \boldsymbol{v}_1) \quad (8.2)$$

A direct computation show shows that $\boldsymbol{u} \cdot \boldsymbol{v}_1 = 6$ and $\boldsymbol{v}_1 \cdot \boldsymbol{v}_1 = 3$ and therefore $6 = 3a_1$. Consequently, $a_1 = 2$. In exactly the same way we obtain $\boldsymbol{u} \cdot \boldsymbol{v}_2 = a_2(\boldsymbol{v}_2 \cdot \boldsymbol{v}_2), \boldsymbol{u} \cdot \boldsymbol{v}_3 = a_3(\boldsymbol{v}_3 \cdot \boldsymbol{v}_3)$. Since $\boldsymbol{u} \cdot \boldsymbol{v}_2 = -3, \boldsymbol{u} \cdot \boldsymbol{v}_3 = -1, \boldsymbol{v}_2 \cdot \boldsymbol{v}_2 = 6$ and $\boldsymbol{v}_3 \cdot \boldsymbol{v}_3 = 2$ we find that $a_2 = -\frac{1}{2}, a_3 = -\frac{1}{2}$.

Example (8.1.1) motivates the following definition:

Definition 8.1. A finite sequence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ of non-zero vectors in $\mathbb{R}^n$ is said to be an ***orthogonal sequence*** if distinct vectors are **orthogonal**, that is, for $i \neq j, \boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$.

If $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is an **orthogonal sequence** of $\mathbb{R}^n$ and a **basis** of a **subspace** W of $\mathbb{R}^n$ then we say that $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is an ***orthogonal basis*** of W.

Orthogonal sequences have properties like the one in Example (8.1.1). In particular, they are **linearly independent**:

Theorem 8.1.1. *Let $\mathcal{S} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be an orthogonal sequence in $\mathbb{R}^n$. Then S is* **linearly independent**

Proof. The proof is by **mathematical induction** on k. Since the vectors in an orthogonal sequence are non-zero, if $k = 1$ (the initial case) then the result is true since a single non-zero vector is **linearly independent**. We now do the inductive case.

So assume that every **orthogonal sequence** of k vectors is **linearly independent** and that $\mathcal{S} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1})$ is an orthogonal sequence. We need to show that $\mathcal{S}$ is linearly independent. Since $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is an orthogonal sequence with k vectors, by the inductive hypothesis, it is linearly independent.

If $\mathcal{S}$ is **linearly dependent** then by **Theorem** (2.4.7) it must be the case that $\boldsymbol{v}_{k+1}$ is a **linear combination** of $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$. So assume that $\boldsymbol{v}_{k+1} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$. We then have

$$\boldsymbol{v}_{k+1} \cdot \boldsymbol{v}_{k+1} = \boldsymbol{v}_{k+1} \cdot (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) =$$

$$c_1(\boldsymbol{v}_{k+1} \cdot \boldsymbol{v}_1) + c_2(\boldsymbol{v}_{k+1} \cdot \boldsymbol{v}_2) + \cdots + c_k(\boldsymbol{v}_{k+1} \cdot \boldsymbol{v}_k) \tag{8.3}$$

by **additivity and the scalar properties** for the **dot product**.

Since $\mathcal{S}$ is an orthogonal sequence, for each $i < k+1$, $\boldsymbol{v}_{k+1} \cdot \boldsymbol{v}_i = 0$. But then from Equation (8.3) it follows that $\boldsymbol{v}_{k+1} \cdot \boldsymbol{v}_{k+1} = 0$ and therefore $\boldsymbol{v}_{k+1} = \boldsymbol{0}_n$, a contradiction to the fact that $\boldsymbol{v}_{k+1} \neq \boldsymbol{0}_n$ and **positive definiteness** of the **dot product**. $\square$

As in **Example** (8.1.1) it is also the case that for an orthogonal sequence $\mathcal{S} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ in $\mathbb{R}^n$ it is easy to compute the **coordinate vector** for $\boldsymbol{u}$ in $Span(\mathcal{S})$ with respect to $\mathcal{S}$. This is the subject of our next theorem.

Theorem 8.1.2. *Let $\mathcal{S} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ be an* **orthogonal sequence** *and $\boldsymbol{u}$ be a vector in $Span(\mathcal{S})$. If $[\boldsymbol{u}]_{\mathcal{S}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ is the* **coordinate vector** *of $\boldsymbol{u}$ with respect to $\mathcal{S}$ $c_i = \frac{\boldsymbol{u} \cdot \boldsymbol{v}_i}{\boldsymbol{v}_i \cdot \boldsymbol{v}_i}$.*

Proof. Assume $[\boldsymbol{u}]_{\mathcal{S}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$. This means that $\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$. We then have

$$\boldsymbol{u} \cdot \boldsymbol{v}_i = (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) \cdot \boldsymbol{v}_i =$$

$$c_1(\boldsymbol{v}_1 \cdot \boldsymbol{v}_i) + c_2(\boldsymbol{v}_2 \cdot \boldsymbol{v}_i) + \cdots + c_i(\boldsymbol{v}_i \cdot \boldsymbol{v}_i) + \ldots c_k(\boldsymbol{v}_k \cdot \boldsymbol{v}_i) \tag{8.4}$$

by the **additivity and scalar properties** of the dot product.

Because $\boldsymbol{v}_j \cdot \boldsymbol{v}_i = 0$ for $j \neq i$, Equation (8.4) reduces to $\boldsymbol{u} \cdot \boldsymbol{v}_i = c_i(\boldsymbol{v}_i \cdot \boldsymbol{v}_i)$. Since $\boldsymbol{v}_i$ is non-zero, $\boldsymbol{v}_i \cdot \boldsymbol{v}_i \neq 0$ and we can deduce that $c_i = \frac{\boldsymbol{u} \cdot \boldsymbol{v}_i}{\boldsymbol{v}_i \cdot \boldsymbol{v}_i}$ as claimed. $\square$

A consequence of Theorem (8.1.2) is that if W is a **subspace** of $\mathbb{R}^n$, $\mathcal{S} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ is a **basis** for W, and $\mathcal{S}$ is an **orthogonal sequence** then the computation of the coordinates of a vector $\boldsymbol{u}$ in W is quite easy. The computation of coordinates is even simpler when the vectors in an **orthogonal sequence** are **unit vectors**. Such sequences are the motivation for the next definition.

Definition 8.2. An **orthogonal sequence** $\mathcal{S}$ consisting of **unit vectors** is called an ***orthonormal sequence***. If W is a **subspace** of $\mathbb{R}^n$, $\mathcal{S}$ is a **basis** for W, and $\mathcal{S}$ is an orthonormal sequence then $\mathcal{S}$ is said to be an ***orthonormal basis for*** W.

Example 8.1.2. 1) The **standard basis** $(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)$ is an orthonormal basis for $\mathbb{R}^n$.

2) If we **normalize** (multiply each vector by the reciprocal of its **norm**) the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ from **Example** (8.1.1) we obtain an orthonormal basis of $\mathbb{R}^3$:

$$\mathcal{S}' = \left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right).$$

Suppose $\mathcal{S} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ is an orthonormal basis of $\mathbb{R}^n$. Since $\mathcal{S}$ is a **basis** we know that the matrix $Q = (\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \dots \ \boldsymbol{v}_n)$ obtained from $\mathcal{S}$ is invertible. However, the **inverse** of this matrix is rather easy to compute, specifically, $Q^{-1} = Q^{Tr}$, the **transpose** of Q.

Example 8.1.3. 1) When $\mathcal{S}$ is the **standard basis**, $(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n)$ then Q is the **identity matrix**, I_n and $I_n = I_n^{-1} = I_n^{Tr}$.

2) For the orthonormal basis $\left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right)$ of $\mathbb{R}^3$ the matrix Q is

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \tag{8.5}$$

Then $Q^{Tr} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$ and $Q^{Tr}Q =$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3 \tag{8.6}$$

We give a name to matrices which arise this way:

Definition 8.3. An $n \times n$ matrix Q whose columns are an **orthonormal basis** for $\mathbb{R}^n$ is said to be an ***orthogonal matrix***. More generally, an $n \times m$ matrix $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_m)$ with $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ an **orthonormal sequence** is referred to as an ***orthonormal matrix***.

The next theorem characterizes orthonormal matrices.

Theorem 8.1.3. *Let $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_m)$ be an $n \times m$ matrix. Then Q is an orthonormal matrix if and only if $Q^{Tr}Q = I_m$.*

Proof. First assume that Q is an orthonormal matrix. Then $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$ if $i \neq j$ and $\boldsymbol{v}_i \cdot \boldsymbol{v}_i = 1$. The $(i,j)-$entry of $Q^{Tr}Q$ is the **dot product** of the i^{th} row of Q^{Tr} with the j^{th} column of Q which is the same thing as the **dot product** of the ith and j^{th} columns of Q and so is equal to $\boldsymbol{v}_i \cdot \boldsymbol{v}_j$. Therefore $Q^{Tr}Q = I_m$ as claimed.

Conversely, assume that $Q^{Tr}Q = I_m$. Now $Q^{Tr} = \begin{pmatrix} \boldsymbol{v}_1^{Tr} \\ \boldsymbol{v}_2^{Tr} \\ \vdots \\ \boldsymbol{v}_m^{Tr} \end{pmatrix}$. Then $\boldsymbol{v}_i^{Tr}\boldsymbol{v}_j = \boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$ if $i \neq j$ and 1 if $i = j$. Thus, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ is an **orthonormal sequence** and Q is an **orthonormal matrix**. □

The theorem can be applied to **orthogonal matrices**.

Theorem 8.1.4. *Let Q be an $n \times n$ matrix. Then the following are equivalent:*

1. *Q is an* **orthogonal matrix**.
2. *$Q^{Tr}Q = I_n$*
3. *The rows of Q are an* **orthonormal basis** *for $\mathbb{R}^n$.*

Proof. Condition 1 is equivalent to condition 2 by Theorem (8.1.3).

We next prove that condition 2 implies 3. Since Q is a square matrix, $Q^{Tr}Q = I_n$ implies that $QQ^{Tr} = I_n$ by **Theorem** (3.4.12). Since $Q = (Q^{Tr})^{Tr}$ it follows that Q^{Tr} is an **orthogonal matrix** and consequently the columns of Q^{Tr} form an

orthonormal basis of $\mathbb{R}^n$. Since the columns of Q^{Tr} are the rows of Q we have established 3.

Finally, condition 3 implies 2. Assume that the rows of Q are an **orthonormal basis** of $\mathbb{R}^n$. Then the columns of Q^{Tr} are an **orthonormal basis** of $\mathbb{R}^n$ and therefore Q^{Tr} is an orthogonal matrix. From the equivalence of 1 and 2, this implies that $(Q^{Tr})^{Tr}Q^{Tr} = QQ^{Tr} = I_n$. By **Theorem** (3.4.12) we have that $Q^{Tr}Q = I_n$. □

In the next theorem we obtain a further characterization of orthonormal matrices.

Theorem 8.1.5. *Let $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_m)$. Then the following are equivalent:*

1. Q is an orthonormal matrix.

2. For every vector $\boldsymbol{v} \in \mathbb{R}^m$, $\parallel Q\boldsymbol{v} \parallel\ =\ \parallel \boldsymbol{v} \parallel$.

3. For every pair of vectors $\boldsymbol{u}, \boldsymbol{w} \in \mathbb{R}^m$, $(Q\boldsymbol{u}) \cdot (Q\boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{w}$.

Proof. 1. implies 2. Assume that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ is an **orthonormal sequence** in $\mathbb{R}^n$. Let $\boldsymbol{u} \in \mathbb{R}^m$. Then $\parallel Q\boldsymbol{u} \parallel^2 = (Q\boldsymbol{u}) \cdot (Q\boldsymbol{u}) = (Q\boldsymbol{u})^{Tr}(Q\boldsymbol{u}) = (\boldsymbol{u}^{Tr}Q^{Tr})(Q\boldsymbol{u}) = \boldsymbol{u}^{Tr}(Q^{Tr}Q)\boldsymbol{u}$ since for the product of two matrices $AB, (AB)^{Tr} = B^{Tr}A^{Tr}$ and by associativity of the matrix product.

By **Theorem** (8.1.3), $Q^{Tr}Q = I_m$ and therefore $\boldsymbol{u}^{Tr}(Q^{Tr}Q)\boldsymbol{u} = \boldsymbol{u}^{Tr}(I_m)\boldsymbol{u} = \boldsymbol{u}^{Tr}\boldsymbol{u} = \boldsymbol{u} \cdot \boldsymbol{u} = \parallel \boldsymbol{u} \parallel^2$.

2. implies 3. We first note that for any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^m$ that

$$\parallel \boldsymbol{x}+\boldsymbol{y} \parallel^2 = (\boldsymbol{x}+\boldsymbol{y}) \cdot (\boldsymbol{x}+\boldsymbol{y}) =$$

$$\boldsymbol{x} \cdot \boldsymbol{x} + \boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{y} \cdot \boldsymbol{x} + \boldsymbol{y} \cdot \boldsymbol{y} = \parallel \boldsymbol{x} \parallel^2 + 2(\boldsymbol{x} \cdot \boldsymbol{y}) + \parallel \boldsymbol{y} \parallel^2 \tag{8.7}$$

So now assume for every vector $\boldsymbol{u} \in \mathbb{R}^m$ that $\parallel Q\boldsymbol{u} \parallel\ =\ \parallel \boldsymbol{u} \parallel$. Of course, we can square both sides to get $\parallel Q\boldsymbol{u} \parallel^2 = \parallel \boldsymbol{u} \parallel^2$. Then this holds for $\boldsymbol{u}, \boldsymbol{w}$ and $\boldsymbol{u}+\boldsymbol{w}$:

$$\parallel Q\boldsymbol{u} \parallel^2 = \parallel \boldsymbol{u} \parallel^2, \parallel Q\boldsymbol{w} \parallel^2 = \parallel \boldsymbol{w} \parallel^2, \parallel Q(\boldsymbol{u}+\boldsymbol{w}) \parallel^2 = \parallel \boldsymbol{u}+\boldsymbol{w} \parallel^2 \tag{8.8}$$

From Equation (8.7) applied to $\boldsymbol{u}+\boldsymbol{w}$ and $Q(\boldsymbol{u}+\boldsymbol{w}) = Q\boldsymbol{u}+Q\boldsymbol{w}$ we get

$$\parallel Q\boldsymbol{u}+Q\boldsymbol{w} \parallel^2 = \parallel Q\boldsymbol{u} \parallel^2 + 2[(Q\boldsymbol{u}) \cdot (Q\boldsymbol{w})] + \parallel Q\boldsymbol{w} \parallel^2 = \parallel \boldsymbol{u}+\boldsymbol{w} \parallel^2 =$$

$$\| \boldsymbol{u} \|^2 + 2(\boldsymbol{u} \cdot \boldsymbol{w}) + \| \boldsymbol{w} \|^2 \tag{8.9}$$

Since $\| Q\boldsymbol{u} \|^2 = \| \boldsymbol{u} \|^2, \| Q\boldsymbol{w} \|^2 = \| \boldsymbol{w} \|^2$ we can cancel these in Equation (8.9) and we get that

$2[(Q\boldsymbol{u}) \cdot (Q\boldsymbol{w})] = 2(\boldsymbol{u} \cdot \boldsymbol{w})$. Dividing by 2 yields $(Q\boldsymbol{u}) \cdot (Q\boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{w}$. Since $\boldsymbol{u}$ and $\boldsymbol{w}$ are arbitrary this establishes condition 3.

3. implies 1. Since we are assuming for any vectors $\boldsymbol{u}, \boldsymbol{w}$ that $(Q\boldsymbol{u}) \cdot (Q\boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{w}$ this applies, in particular, to $\boldsymbol{u} = \boldsymbol{e}_i^m$ and $\boldsymbol{w} = \boldsymbol{e}_j^m$ the i^{th} and j^{th} **standard basis vectors** of $\mathbb{R}^m$. Note that $Q\boldsymbol{e}_i^m = \boldsymbol{v}_i$ is the i^{th} column of Q and $Q\boldsymbol{e}_j^m = \boldsymbol{v}_j$ is the j^{th} column of Q. Then $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = (Q\boldsymbol{e}_i^m) \cdot (Q\boldsymbol{e}_j^m) = \boldsymbol{e}_i^m \cdot \boldsymbol{e}_j^m = 1$ if $i = j$ and 0 if $i \neq j$. Therefore the columns of Q are unit vectors and they are mutually **orthogonal**, that is, $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$ is an **orthonormal sequence** in $\mathbb{R}^n$. □

We now introduce **linear transformations** which behave nicely with respect to the **dot product**.

Definition 8.4. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a **linear transformation**. We say that T ***preserves distances*** if $dist(T(\boldsymbol{v}), T(\boldsymbol{w})) = dist(\boldsymbol{v}, \boldsymbol{w})$ for all $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^m$.

The following theorem characterizes distance preserving **linear transformations**.

Theorem 8.1.6. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a* **linear transformation** *and assume the* **standard matrix** *of T is Q. Then T preserves distances if and only if Q is an* **orthonormal matrix**.

Proof. Assume T preserves distances. Then for all $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ we have

$$dist(T(\boldsymbol{v}), T(\boldsymbol{w})) = dist(\boldsymbol{v}, \boldsymbol{w}).$$

In particular, this holds for $\boldsymbol{w} = \boldsymbol{0}_n$. Therefore, for all $\boldsymbol{v} \in \mathbb{R}^n$ we have

$$\| T(\boldsymbol{v}) \| = dist(T(\boldsymbol{v}), \boldsymbol{0}_n) = dist(\boldsymbol{v}, \boldsymbol{0}_n) = \| \boldsymbol{v} \| .$$

By the **definition of the standard matrix** it follows that for all $\boldsymbol{v} \in \mathbb{R}^n$

$$\| Q\boldsymbol{v} \| = \| \boldsymbol{v} \| .$$

By Theorem (8.1.5) Q is an **orthonormal matrix**.

Conversely, assume Q is an orthonormal matrix. Let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$. By the definition of distance between two vectors we have

$$dist(Q\boldsymbol{v}, Q\boldsymbol{w}) = \parallel Q\boldsymbol{v} - Q\boldsymbol{w} \parallel$$

By the distributive property for matrix multiplication we can conclude that

$$\parallel Q\boldsymbol{v} - Q\boldsymbol{w} \parallel = \parallel Q(\boldsymbol{v} - \boldsymbol{w}) \parallel$$

Now by **Theorem** (8.1.5) it follows that

$$\parallel Q(\boldsymbol{v} - \boldsymbol{w}) \parallel = \parallel \boldsymbol{v} - \boldsymbol{w} \parallel = dist(\boldsymbol{v}, \boldsymbol{w}).$$

We have therefore shown that $dist(Q\boldsymbol{v}, Q\boldsymbol{w}) = dist(\boldsymbol{v}, \boldsymbol{w})$. Since $T(\boldsymbol{v}) = Q\boldsymbol{v}, T(\boldsymbol{w}) = Q\boldsymbol{w}$ this completes the proof. □

Theorem (8.1.6) motivates the following definition:

Definition 8.5. A **linear operator** $T : \mathbb{R}^n \to \mathbb{R}^n$ is an ***orthogonal transformation*** if it **preserves distances**.

We now consider **linear transformations** which preserve angles.

Definition 8.6. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation**. We say that T ***preserves angles*** if for every pair of vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ $\angle(T(\boldsymbol{v}), T(\boldsymbol{w})) = \angle(\boldsymbol{v}, \boldsymbol{w})$.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** and the **standard matrix** of T is an **orthonormal matrix** then T preserve angles. We state this as a theorem but leave it as a **challenge exercise**:

Theorem 8.1.7. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a* **linear transformation** *with* **standard matrix** *$Q = (\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \dots \ \boldsymbol{v}_n)$ and assume that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ is an* **orthonormal sequence** *in $\mathbb{R}^n$. Then T preserves angles.*

Theorem (8.1.7) is not a characterization theorem since there are other **linear transfor–mations** which preserve angles. This is the subject of our last example.

Example 8.1.4. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be scalar multiplication by the scalar $c \neq 0$ so that the matrix of T (with respect to any basis) is cI_n. Then T **preserves angles**. The matrix cI_n is not an **orthogonal matrix** when $c \neq \pm 1$. Moreover, the composition of **angle preserving linear transformations** is angle preserving (left as a **challenge exercise**). Therefore, multiplying a scalar matrix cI_n $(c \neq \pm 1)$ by an **orthonormal matrix** gives the matrix of an **angle preserving linear transformation** which is also not **orthonormal**.

What You Can Now do

1. Given a sequence of vectors $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ from $\mathbb{R}^n$ determine if it is an **orthogonal sequence**.

2. Given a sequence of vectors $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ from $\mathbb{R}^n$ determine if it is an **orthonormal sequence**.

3. Given an **orthogonal sequence** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ and a vector $\boldsymbol{u}$ in $Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ compute the **coordinate vector** of $\boldsymbol{u}$ with respect to $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$.

4. Determine if an $n \times n$ matrix Q is an **orthogonal matrix**.

Method (How To Do It)

Method 8.1.1. Given a sequence of vectors $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ from $\mathbb{R}^n$ determine if it is an **orthogonal sequence**.

Compute all the dot products $\boldsymbol{w}_i \cdot \boldsymbol{w}_j$ for $i \neq j$. If they are all zero then the sequence is an **orthogonal sequence**, otherwise it is not.

Example 8.1.5. Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 2 \\ 3 \\ -2 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 16 \\ -10 \\ 1 \end{pmatrix}$. Determine if $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an **orthogonal sequence**.

$$\boldsymbol{w}_1 \cdot \boldsymbol{w}_2 = (1)(2)+(2)(3)+(4)(-2) = 0, \boldsymbol{w}_1 \cdot \boldsymbol{w}_3 = (1)(16)+(2)(-10)+(4)(1) = 0,$$

$$\boldsymbol{w}_2 \cdot \boldsymbol{w}_3 = (2)(16) + (3)(-10) + (-2)(1) = 0.$$

It is an **orthogonal sequence**.

Example 8.1.6. Let $\boldsymbol{w}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix}2\\-5\\3\end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix}-4\\1\\3\end{pmatrix}$. Determine if $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an **orthogonal sequence**.

$$\boldsymbol{w}_1 \cdot \boldsymbol{w}_2 = (1)(2)+(1)(-5)+(1)(3) = 0, \boldsymbol{w}_1 \cdot \boldsymbol{w}_3 = (1)(-4)+(1)(1)+(1)(3) = 0,$$

$$\boldsymbol{w}_2 \cdot \boldsymbol{w}_3 = (2)(-4)+(-5)(1)+(3)(3) = -4 \neq 0.$$

The sequence $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is not an **orthogonal sequence**.

Example 8.1.7. Let $\boldsymbol{w}_1 = \begin{pmatrix}1\\2\\2\\0\end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix}2\\-1\\0\\2\end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix}2\\4\\-5\\0\end{pmatrix}, \boldsymbol{w}_4 = \begin{pmatrix}4\\-2\\0\\-5\end{pmatrix}$. Determine if $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4)$ is an **orthogonal sequence**.

$\boldsymbol{w}_1 \cdot \boldsymbol{w}_2 = (1)(2)+(2)(-1)+(2)(0)+(0)(1) = 0,$

$\boldsymbol{w}_1 \cdot \boldsymbol{w}_3 = (1)(2)+(2)(4)+(2)(-5)+(0)(1) = 0,$

$\boldsymbol{w}_1 \cdot \boldsymbol{w}_4 = (1)(4)+(2)(-2)+(2)(0)+(0)(-5) = 0,$

$\boldsymbol{w}_2 \cdot \boldsymbol{w}_3 = (2)(2)+(-1)(4)+(0)(-5)+(2)(0) = 0,$

$\boldsymbol{w}_2 \cdot \boldsymbol{w}_4 = (2)(4)+(-1)(-2)+(0)(0)+(2)(-5) = 0,$

$\boldsymbol{w}_3 \cdot \boldsymbol{w}_4 = (2)(4)+(4)(-2)+(-5)(0)+(0)(-5) = 0.$

The sequence is an **orthogonal sequence**.

Method 8.1.2. Given a sequence of vectors $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ from $\mathbb{R}^n$ determine if it is an **orthonormal sequence**.

First check to see if every vector is a **unit vector** by computing $\boldsymbol{w}_i \cdot \boldsymbol{w}_i$. If for some i this is not 1 then stop, the set is not an **orthonormal sequence**.

If for each i, $\boldsymbol{w}_i \cdot \boldsymbol{w}_i = 1$ then check to see if the sequence is an **orthogonal sequence**. If so, the the sequence it is an **orthonormal sequence**, otherwise it is not.

Example 8.1.8. Let $\boldsymbol{w}_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$. Determine if $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an **orthonormal sequence**.

$$\boldsymbol{w}_1 \cdot \boldsymbol{w}_1 = (\frac{1}{\sqrt{3}})^2 + (\frac{1}{\sqrt{3}})^2 + (\frac{1}{\sqrt{3}})^2 = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$$

$$\boldsymbol{w}_2 \cdot \boldsymbol{w}_2 = (\frac{2}{\sqrt{6}})^2 + (-\frac{1}{\sqrt{6}})^2 + (-\frac{1}{\sqrt{6}})^2 = \frac{4}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

$$\boldsymbol{w}_3 \cdot \boldsymbol{w}_3 = 0^2 + (\frac{1}{\sqrt{2}})^2 + (-\frac{1}{\sqrt{2}})^2 = 0 + \frac{1}{2} + \frac{1}{2} = 1.$$

Thus, $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ are all **unit vectors**.

$$\boldsymbol{w}_1 \cdot \boldsymbol{w}_2 = (\frac{1}{\sqrt{3}})(\frac{2}{\sqrt{6}}) + (\frac{1}{\sqrt{3}})(-\frac{1}{\sqrt{6}}) + (\frac{1}{\sqrt{3}})(-\frac{1}{\sqrt{6}}) = \frac{2}{\sqrt{18}} - \frac{1}{\sqrt{18}} - \frac{1}{\sqrt{18}} = 0,$$

$$\boldsymbol{w}_1 \cdot \boldsymbol{w}_3 = (\frac{1}{\sqrt{3}})(0) + (\frac{1}{\sqrt{3}})(\frac{1}{\sqrt{2}}) + (\frac{1}{\sqrt{3}})(-\frac{1}{\sqrt{2}}) = 0 + \frac{1}{\sqrt{6}} - \frac{1}{\sqrt{6}} = 0,$$

$$\boldsymbol{w}_2 \cdot \boldsymbol{w}_3 = (\frac{2}{\sqrt{6}})(0) + (-\frac{1}{\sqrt{6}})(\frac{1}{\sqrt{2}}) + (-\frac{1}{\sqrt{6}})(-\frac{1}{\sqrt{2}}) = 0 + \frac{1}{12} - \frac{1}{12} = 0.$$

We can conclude that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an **orthonormal sequence**.

Method 8.1.3. Given an **orthogonal sequence** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ and a vector $\boldsymbol{u}$ in $Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ compute the **coordinate vector** of $\boldsymbol{u}$ with respect to $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$.

Compute $c_i = \frac{\boldsymbol{u} \cdot \boldsymbol{w}_i}{\boldsymbol{w}_i \cdot \boldsymbol{w}_i}$ for $i = 1, 2, \dots, k$. Then $\boldsymbol{u} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \cdots + c_k\boldsymbol{w}_k$.

Example 8.1.9. Let $\boldsymbol{w}_1 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. Verify that this is an **orthogonal sequence** and a **basis** for $\mathbb{R}^3$. Then find the **coordinate vector** of $\boldsymbol{u} = \begin{pmatrix} 8 \\ 3 \\ 1 \end{pmatrix}$.

$\boldsymbol{w}_1 \cdot \boldsymbol{w}_2 = (2)(1) + (1)(-1) + (-1)(1) = 0,$

$\boldsymbol{w}_1 \cdot \boldsymbol{w}_3 = (2)(0) + (1)(1) + (-1)(1) = 0,$

$\boldsymbol{w}_2 \cdot \boldsymbol{w}_3 = (1)(0) + (-1)(1) + (1)(1) = 0.$

Thus, $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an **orthogonal sequence**. By **Theorem** (8.1.1) $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent**. By **Theorem** (2.4.4) it is a **basis** for $\mathbb{R}^3$.

$$\frac{\boldsymbol{u} \cdot \boldsymbol{w}_1}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} = \frac{18}{6} = 3, \frac{\boldsymbol{u} \cdot \boldsymbol{w}_2}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2} = \frac{6}{3} = 2, \frac{\boldsymbol{u} \cdot \boldsymbol{w}_3}{\boldsymbol{w}_3 \cdot \boldsymbol{w}_3} = \frac{4}{2} = 2.$$

$$\begin{pmatrix} 8 \\ 3 \\ 1 \end{pmatrix} = 3 \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Example 8.1.10. Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 1 \\ 1 \\ -3 \\ 1 \end{pmatrix}$. Prove $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an **orthogonal basis** for the **subspace**

$$W = \{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} : x_1 + x_2 + x_3 + x_4 = 0\}$$

of $\mathbb{R}^4$. Then find the **coordinate vector** of $\boldsymbol{u} = \begin{pmatrix} 3 \\ 8 \\ -6 \\ -5 \end{pmatrix}$.

We begin by showing that $\mathcal{B}$ is an **orthogonal sequence**:

$$\boldsymbol{w}_1 \cdot \boldsymbol{w}_2 = (1)(1) + (0)(-2) + (0)(0) + (-1)(1) = 0,$$

$$\boldsymbol{w}_1 \cdot \boldsymbol{w}_3 = (1)(1) + (0)(1) + (0)(-3) + (1)(1) = 0,$$

$$\boldsymbol{w}_2 \cdot \boldsymbol{w}_3 = (1)(1) + (-2)(1) + (0)(-3) + (1)((1) = 0.$$

So, $\mathcal{B}$ is an **orthogonal sequence** and therefore **linearly independent**. Clearly each vector belongs to W (add up the entries). We claim that W has **dimension** three. First note that W is a proper **subspace** of $\mathbb{R}^4$ (the **standard basis vectors** are not in W). This implies that $dim(W) \leq 3$. Since $\mathcal{B}$ is a **linearly independent** sequence in W with three vectors it must be that $dim(W) \geq 3$. Therefore $dim(W) = 3$ and by **Theorem** (2.4.4) $\mathcal{B}$ is a **basis** for W.

The vector $\boldsymbol{u}$ belongs to W since the sum of its components is zero. We compute $\frac{\boldsymbol{u} \cdot \boldsymbol{w}_i}{\boldsymbol{w}_i \cdot \boldsymbol{w}_i}$ for $i = 1, 2, 3$:

$$\frac{\boldsymbol{u} \cdot \boldsymbol{w}_1}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} = \frac{8}{2} = 4, \frac{\boldsymbol{u} \cdot \boldsymbol{w}_2}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2} = \frac{-18}{6} = -3, \frac{\boldsymbol{u} \cdot \boldsymbol{w}_3}{\boldsymbol{w}_3 \cdot \boldsymbol{w}_3} = \frac{24}{12} = 2.$$

The **coordinate vector** of $\boldsymbol{u}$ with respect to $\mathcal{B}$ is $\begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix}$.

Method 8.1.4. Determine if an $n \times n$ matrix Q is an **orthogonal matrix**.

Multiply Q^{Tr} by Q. If the result is the **identity matrix** then Q is an **orthogonal matrix**, otherwise it is not.

Example 8.1.11. Determine if the matrix $Q = \begin{pmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{45}} \\ \frac{2}{3} & -\frac{1}{\sqrt{5}} & \frac{4}{\sqrt{45}} \\ \frac{2}{3} & 0 & -\frac{5}{\sqrt{45}} \end{pmatrix}$ is an **orthogonal matrix**.

$$Q^T Q = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{45}} & \frac{4}{\sqrt{45}} & -\frac{5}{\sqrt{45}} \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{45}} \\ \frac{2}{3} & -\frac{1}{\sqrt{5}} & \frac{4}{\sqrt{45}} \\ \frac{2}{3} & 0 & -\frac{5}{\sqrt{45}} \end{pmatrix} = I_3.$$

Q is an **orthogonal matrix**.

Example 8.1.12. Determine if the matrix $Q = \begin{pmatrix} \frac{1}{\sqrt{5}} & -2 \\ \frac{2}{\sqrt{5}} & 1 \end{pmatrix}$ is an **orthogonal matrix**.

$$Q^T Q = \begin{pmatrix} \frac{1}{\sqrt{25}} & \frac{2}{\sqrt{5}} \\ -2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -2 \\ \frac{2}{\sqrt{5}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$$

The matrix Q is not an **orthogonal matrix**. However, because the product $Q^{Tr}Q$ is a **diagonal matrix** the columns of the matrix are **orthogonal**, however, the second column is not a **unit vector**.

Exercises

In exercises 1-5 determine if the given sequence of vectors is an **orthogonal sequence**. See **Method** (8.1.1).

1. $\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} -6 \\ 2 \end{pmatrix}\right)$.

2. $\left(\begin{pmatrix} 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix}\right)$.

3. $\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \\ -5 \end{pmatrix}\right)$.

4. $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ -3 \\ 3 \end{pmatrix}\right)$.

5. $\left(\begin{pmatrix} 1 \\ 2 \\ -2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 \\ -2 \\ 3 \\ 2 \end{pmatrix}\right)$.

In exercises 6 - 12 determine if the given sequence of vectors is an **orthonormal basis** for the appropriate $\mathbb{R}^n$. See **Method** (8.1.2). Note if the vectors belong to $\mathbb{R}^n$ then the length of the sequence must be n.

6. $\left(\begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}\right)$

7. $\left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$

8. $\left(\begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}\right)$

9. $\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \end{pmatrix}\right)$

10. $\left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}\right)$

11. $\left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}\right)$

12. $\left(\begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{5}{\sqrt{30}} \\ -\frac{1}{\sqrt{30}} \\ -\frac{2}{\sqrt{30}} \end{pmatrix}\right)$

In exercises 13 - 15 find the **coordinate vector** of the given vector $\boldsymbol{u}$ with respect to the given **orthogonal basis**, $\mathcal{O}$, for $\mathbb{R}^n$. See **Method** (8.1.3).

13. $\mathcal{O} = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$.

14. $\mathcal{O} = \left(\begin{pmatrix} 3 \\ 5 \end{pmatrix}, \begin{pmatrix} -5 \\ 3 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 1 \\ 13 \end{pmatrix}$.

15. $\mathcal{O} = \left(\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -4 \\ -5 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 9 \\ -8 \\ 1 \end{pmatrix}$.

16. a) Prove that the sequence $\mathcal{O} = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 3 \\ 1 \end{pmatrix}\right)$ is an **orthogonal basis** for the **subspace** $W = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}^{\perp}$. See **Method** (8.1.1), prove $dim(W) = 3$ and use **Theorem** (2.4.4).

b) Demonstrate that the vector $\boldsymbol{u} = \begin{pmatrix} 1 \\ 1 \\ 3 \\ 5 \end{pmatrix}$ is in W and find the **coordinate vector** of $\boldsymbol{u}$ with respect to $\mathcal{O}$. See **Method** (8.1.3).

17. a) Prove that the sequence $\mathcal{O} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ -3 \\ 4 \end{pmatrix} \right)$ is an **orthogonal basis** for the **subspace** $W = Span\left(\begin{pmatrix} 1 \\ -4 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right)^{\perp}$ of $\mathbb{R}^4$. See **Method** (8.1.1). Prove $dim(W) = 2$ and use **Theorem** (2.4.4).

b) Demonstrate that the vector $\boldsymbol{u} = \begin{pmatrix} 3 \\ 0 \\ -1 \\ 6 \end{pmatrix}$ is in W and find the **coordinate vector** of $\boldsymbol{u}$ with respect to $\mathcal{O}$. See **Method** (8.1.3).

18. a) Prove that $\mathcal{O} = \left(\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ 2 \end{pmatrix} \right)$ is an **orthogonal basis** for the **subspace** $W = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}^{\perp}$ of $\mathbb{R}^3$. See **Method** (8.1.1). Prove $dim(W) = 2$ and use **Theorem** (2.4.4).

b) Let $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \right)$. Prove that $\mathcal{B}$ is a **basis** for W and find the **change of basis matrix** $P_{\mathcal{B} \to \mathcal{O}}$.

In exercises 19 - 25 determine if the given matrix Q is an **orthogonal matrix**. See **Method** (8.1.4).

19. $Q = \begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix}$

20. $Q = \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{pmatrix}$

21. $Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}$

22. $Q = \begin{pmatrix} \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{1}{\sqrt{2}} \\ -\frac{2}{3} & \frac{1}{\sqrt{2}} \end{pmatrix}$

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

24. $Q = \begin{pmatrix} \frac{1}{\sqrt{11}} & \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{110}} \\ \frac{1}{\sqrt{11}} & 0 & \frac{10}{\sqrt{110}} \\ \frac{3}{\sqrt{11}} & -\frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{110}} \end{pmatrix}$

25. $Q = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{14}} & \frac{5}{\sqrt{42}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{14}} & -\frac{4}{\sqrt{42}} \\ -\frac{1}{\sqrt{3}} & \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{42}} \end{pmatrix}$

In exercises 26 - 31 answer true or false and give an explanation.

26. Every **linearly independent sequence** in $\mathbb{R}^n$ is an **orthogonal sequence**.

27. Every **orthogonal sequence** in $\mathbb{R}^n$ is **linearly independent**.

28. If the rows of a matrix A are mutually **orthogonal** and an **elementary row operation** is applied then the rows of the matrix obtained are mutually **orthogonal**.

29. An **orthogonal matrix** is **invertible**.

30. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ is an **orthogonal sequence** in $\mathbb{R}^n$ and $c_1, c_2, \dots, c_k$ are non-zero scalars then $(c_1\boldsymbol{v}_1, c_2\boldsymbol{v}_2, \dots, c_k\boldsymbol{v}_k)$ is an **orthogonal sequence**.

31. If the rows of a matrix A are mutually **orthogonal** and an exchange operation is applied then the rows of the matrix obtained are mutually **orthogonal**.

Challenge Exercises (Problems)

1. Let $S : \mathbb{R}^n \to \mathbb{R}^m$ be a **linear transformation** and assume that the **standard matrix** of S is **orthonormal**. Prove that T **preserves angles**.

2. a) Let $S : \mathbb{R}^m \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^p$ be a **linear transformations**. Assume that S and T **preserve distances**. Prove that $T \circ S : \mathbb{R}^m \to \mathbb{R}^p$ **preserves distances**.

b) Let A be an $n \times m$ **orthonormal matrix** and B a $p \times n$ **orthonormal matrix**. Prove that BA is a $p \times m$ **orthonormal matrix**.

3. Let $S : \mathbb{R}^m \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^p$ be a **linear transformations**. Assume that S and T **preserve angles**. Prove that $T \circ S : \mathbb{R}^m \to \mathbb{R}^p$ **preserves angles**.

4. Let Q_1, Q_2 be two **orthogonal** $n \times n$ matrices. Prove that Q_1Q_2 is an **orthogonal**

matrix. (Hint: Prove that Q_1Q_2 is **invertible** and that $(Q_1Q_2)^{-1} = (Q_1Q_2)^{Tr}$.)

5. Assume that Q is an $n \times n$ **orthogonal matrix** and Q is **symmetric**. Prove that $Q^2 = I_n$.

6. Let $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_m)$ be a sequence of vectors from $\mathbb{R}^n$ and let $Q = (\boldsymbol{w}_1\ \boldsymbol{w}_2\ \dots\ \boldsymbol{w}_m)$. Prove that $\mathcal{B}$ is an **orthogonal sequence** if and only if $Q^{Tr}Q$ is a **diagonal matrix** with non-zero diagonal entries.

7. Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be two **orthonormal bases** of $\mathbb{R}^n$. Prove that the **change of basis matrix** $P_{\mathcal{O}_1 \to \mathcal{O}_2}$ is an **orthogonal matrix**.

Quiz Solutions

1. In a), b), and c) the vectors are **orthogonal**. In d) the vectors are not **orthogonal**. Not right, see **definition of orthogonal vectors**.

2. a) $\begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$ b) $\begin{pmatrix} \frac{1}{5} \\ -\frac{2}{5} \\ \frac{4}{5} \\ -\frac{2}{5} \end{pmatrix}$ c) $\begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \end{pmatrix}$ d) $\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$

Not right, see **definition of the length of a vector**, **definition of a unit vector**, and **Remark** (2.11).

3. The **reduced echelon form** of the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$ is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ which proves the sequence of vectors $\mathcal{S} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \right)$ is **linearly independent**.

Not right, see **Method** (2.4.1).

b) The **reduced echelon form** of the matrix $\left(\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 1 & 1 & -1 & 4 \\ 1 & -1 & 1 & 2 \\ 1 & -1 & -1 & 0 \end{array}\right)$ is $\left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$ which proves that $\boldsymbol{u} \in Span(\mathcal{S})$ and that the **coordinate vector** of $\boldsymbol{u}$ with respect to $\mathcal{S}$ is $[\boldsymbol{u}]_{\mathcal{S}} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$. Not right, see **Method** (2.3.1) and **Method** (2.5.4).

8.2. The Gram-Schmidt Process and QR-Factorization

In this section we give a method for constructing an **orthonormal basis** for a **subspace** W of $\mathbb{R}^n$ when we are given a **basis** for W. This is known as the Gram-Schmidt process. We then use this to express an $n \times m$ matrix A with **linearly independent** columns as a product QR with Q an $n \times m$ **orthonormal matrix** and R is an $m \times m$ **invertible matrix**.

Am I Ready for This Material

The following are important to an understanding of this section

dot product

orthogonal vectors

the length (norm, magnitude) of a vector

unit vector

orthogonal sequence

orthogonal basis for a subspace of $\mathbb{R}^n$

orthonormal sequence

orthonormal basis for a subspace of $\mathbb{R}^n$

orthonormal matrix

orthogonal matrix

orthogonal complement of a subspace

projection of a vector onto the span of a given vector

projection of a vector orthogonal to a given vector

Quiz

1. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$. Find a non-zero vector $\boldsymbol{v}_3$ such that $\boldsymbol{v}_1 \perp \boldsymbol{v}_3 \perp \boldsymbol{v}_2$.

2. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 0 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 1 \end{pmatrix}$ and set $V = Span(\boldsymbol{v}_1, \boldsymbol{v}_2)$. Find a **basis** for the **orthogonal complement** to V.

3. Let $\boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$. Find vectors $\boldsymbol{x}, \boldsymbol{y}$ such that $\boldsymbol{u} = \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} \in Span(\boldsymbol{v})$ and $\boldsymbol{y}$ is **orthogonal** to $\boldsymbol{v}$.

4. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$. Find vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}$ such that

$$\boldsymbol{u} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{y}, \boldsymbol{x}_1 \in Span(\boldsymbol{v}_1), \boldsymbol{x}_2 \in Span(\boldsymbol{v}_2), \boldsymbol{v}_1 \perp \boldsymbol{y} \perp \boldsymbol{v}_2.$$

Quiz Solutions

New Concepts

In this section we introduce one new concept:

QR-factorization of a matrix

Theory (Why It Works)

We now come to a method which will produce an **orthonormal basis** for a **subspace** W of $\mathbb{R}^n$ from a given **basis** for W. This procedure is known as the **Gram-Schmidt Process**.

Gram-Schmidt Process

Assume that W is a **subspace** of $\mathbb{R}^n$ and that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ is a **basis** for W. We shall first define a sequence of vectors $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$ recursively which will be an **orthogonal sequence** and additionally have the property that for each $k, 1 \leq k \leq m, Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$. We then obtain an **orthonormal sequence**, $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m)$, by **normalizing** each vector of the vectors $\boldsymbol{x}_i$, that is, we set

$$\boldsymbol{v}_1 = \frac{1}{\| \boldsymbol{x}_1 \|}\boldsymbol{x}_1, \boldsymbol{v}_2 = \frac{1}{\| \boldsymbol{x}_2 \|}\boldsymbol{x}_2, \ldots, \boldsymbol{v}_m = \frac{1}{\| \boldsymbol{x}_m \|}\boldsymbol{x}_m$$

To say that we **define the sequence recursively** means that we will initially define $\boldsymbol{x}_1$. Then, assuming that we have defined $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$ with the required properties, we will define $\boldsymbol{x}_{k+1}$ such that $\boldsymbol{x}_{k+1}$ is **orthogonal** to $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$. It will be a consequence of our definition that $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{k+1}) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{k+1})$. Since the sequence $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{k+1})$ is **linearly independent** it will then follow from **Theorem** (2.4.4) that the sequence $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{k+1})$ is **linearly independent**. In particular, $\boldsymbol{x}_{k+1} \neq \boldsymbol{0}_n$.

The Definition of $\boldsymbol{x}_1$

We begin with the definition of $\boldsymbol{x}_1$ which we set equal to $\boldsymbol{w}_1$.

The Recursion

To get a sense of what we are doing, we first show how to define of $\boldsymbol{x}_2$ in terms of $\boldsymbol{w}_2$ and $\boldsymbol{x}_1$ and then $\boldsymbol{x}_3$ in terms of $\boldsymbol{x}_1, \boldsymbol{x}_2$ and $\boldsymbol{w}_3$ before doing the arbitrary case.

Defining $\boldsymbol{x}_2$

The idea is to find a **linear combination** $\boldsymbol{x}_2$ of $\boldsymbol{w}_2$ and $\boldsymbol{x}_1$ which is **orthogonal** to $\boldsymbol{x}_1$. The vector $\boldsymbol{x}_2$ will be obtained by adding a suitable multiple of $\boldsymbol{x}_1$ to $\boldsymbol{w}_2$. Consequently, we will have that $Span(\boldsymbol{x}_1, \boldsymbol{x}_2) = Span(\boldsymbol{x}_1, \boldsymbol{w}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$.

Rather than just write down a formula we compute the necessary scalar: Assume that $\boldsymbol{x}_2 = \boldsymbol{w}_2 + a\boldsymbol{x}_1$ and that $\boldsymbol{x}_2 \cdot \boldsymbol{x}_1 = 0$. Then

$$0 = \boldsymbol{x}_2 \cdot \boldsymbol{x}_1 = (\boldsymbol{w}_2 + a\boldsymbol{x}_1) \cdot \boldsymbol{x}_1 = \boldsymbol{w}_2 \cdot \boldsymbol{x}_1 + a(\boldsymbol{x}_1 \cdot \boldsymbol{x}_1) \tag{8.10}$$

We solve for a and obtain

$$a = -\frac{\boldsymbol{w}_2 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1} \tag{8.11}$$

Using the value of a obtained in Equation (8.11) we set $\boldsymbol{x}_2 = \boldsymbol{w}_2 - \frac{\boldsymbol{w}_2 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1$.

We check that $\boldsymbol{x}_2 \cdot \boldsymbol{x}_1$ is, indeed, zero:

$\boldsymbol{x}_2 \cdot \boldsymbol{x}_1 = [\boldsymbol{w}_2 - \frac{\boldsymbol{w}_2 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1] \cdot \boldsymbol{x}_1 = \boldsymbol{w}_2 \cdot \boldsymbol{x}_1 - (\frac{\boldsymbol{w}_2 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1})(\boldsymbol{x}_1 \cdot \boldsymbol{x}_1) = \boldsymbol{w}_2 \cdot \boldsymbol{x}_1 - \boldsymbol{w}_2 \cdot \boldsymbol{x}_1 = 0$

making use of the **additivity and scalar properties** of the **dot product**.

Defining $\boldsymbol{x}_3$

Now that we have defined $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ we find a vector $\boldsymbol{x}_3$ which will be a **linear combination** of the form $\boldsymbol{x}_3 = \boldsymbol{w}_3 + a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2$. We want to determine a_1, a_2 such that $\boldsymbol{x}_3$ is **orthogonal** to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Since $\boldsymbol{x}_3$ and $\boldsymbol{x}_1$ are to be **orthogonal** we must have

$$0 = \boldsymbol{x}_3 \cdot \boldsymbol{x}_1 = (\boldsymbol{w}_3 + a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2) \cdot \boldsymbol{x}_1 = \boldsymbol{w}_1 \cdot \boldsymbol{x}_1 + a_1(\boldsymbol{x}_1 \cdot \boldsymbol{x}_1) + a_2(\boldsymbol{x}_2 \cdot \boldsymbol{x}_1) \tag{8.12}$$

The second equality holds by the **additivity and the scalar properties** of the **dot product**.

Because $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are **orthogonal** we get

$$0 = \boldsymbol{w}_3 \cdot \boldsymbol{x}_1 + a_1(\boldsymbol{x}_1 \cdot \boldsymbol{x}_1), a_1 = -\frac{\boldsymbol{w}_3 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1} \tag{8.13}$$

In an entirely analogous way, using the fact that $\boldsymbol{x}_3$ and $\boldsymbol{x}_2$ are **orthogonal** we obtain

$$a_2 = -\frac{\boldsymbol{w}_3 \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2} \tag{8.14}$$

Thus, $\boldsymbol{x}_3 = \boldsymbol{w}_3 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2}\boldsymbol{x}_2$

We leave it as a **challenge exercise** to show that $\boldsymbol{x}_3$ defined this way is **orthogonal** to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Since $\boldsymbol{x}_3$ is obtained from $\boldsymbol{w}_3$ by adding a **linear combination** of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to $\boldsymbol{w}_3$ we will have that $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{w}_3)$. Since $Span(\boldsymbol{x}_1, \boldsymbol{x}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$ it then follows that $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$. Since $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent**, $dim(Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)) = 3$. It then must be the case that $\boldsymbol{x}_3 \neq \mathbf{0}_n$ and $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ is **linearly independent** by **Theorem** (2.4.4).

The General Recursive Case

We now do the general case. So assume that $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ have been defined with $k < m$ satisfying i) $\boldsymbol{x}_i \cdot \boldsymbol{x}_j = 0$ for $i \neq j$; and $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$.

Then set

$$\boldsymbol{x}_{k+1} = \boldsymbol{w}_{k+1} - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1 - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2}\boldsymbol{x}_2 - \cdots - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_k}{\boldsymbol{x}_k \cdot \boldsymbol{x}_k}\boldsymbol{x}_k \tag{8.15}$$

We show that $\boldsymbol{x}_{k+1} \cdot \boldsymbol{x}_i = 0$ for all $i, 1 \leq i \leq k$.

$$\boldsymbol{x}_{k+1} \cdot \boldsymbol{x}_i = [\boldsymbol{w}_{k+1} - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1 - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2}\boldsymbol{x}_2 - \cdots - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_k}{\boldsymbol{x}_k \cdot \boldsymbol{x}_k}\boldsymbol{x}_k] \cdot \boldsymbol{x}_i =$$

$$\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_i - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}(\boldsymbol{x}_1 \cdot \boldsymbol{x}_i) - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2}(\boldsymbol{x}_2 \cdot \boldsymbol{x}_i)$$

$$- \cdots - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_k}{\boldsymbol{x}_k \cdot \boldsymbol{x}_k}(\boldsymbol{x}_k \cdot \boldsymbol{x}_i) \tag{8.16}$$

by the **additivity and scalar properties** of the **dot product**. Since $\boldsymbol{x}_j \cdot \boldsymbol{x}_i = 0$ for $i \neq j$ Equation (8.16) becomes

$$\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_i - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_i}{\boldsymbol{x}_i \cdot \boldsymbol{x}_i}(\boldsymbol{x}_i \cdot \boldsymbol{x}_i) = \boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_i - \boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_i = 0 \tag{8.17}$$

So, indeed, $\boldsymbol{x}_{k+1}$ as defined is **orthogonal** to $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$.

Because $\boldsymbol{x}_{k+1}$ is obtained from $\boldsymbol{w}_{k+1}$ by adding a **linear combination** of $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k)$ to $\boldsymbol{w}_{k+1}$ it then follows that $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \boldsymbol{x}_{k+1}) = Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \boldsymbol{w}_{k+1})$. Since $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ we can conclude that, $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \boldsymbol{x}_{k+1}) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k, \boldsymbol{w}_{k+1})$. By **Theorem** (2.4.4) we can conclude that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k+1})$ is **linearly independent**. In particular, this implies that $\boldsymbol{x}_{k+1} \neq \mathbf{0}_n$.

Now **normalize** each $\boldsymbol{x}_i$ to obtain $\boldsymbol{v}_i$:

$$\boldsymbol{v}_i = \frac{1}{\parallel \boldsymbol{x}_i \parallel}\boldsymbol{x}_i, i = 1, 2, \ldots, m$$

Since each $\boldsymbol{v}_i$ is obtained from $\boldsymbol{x}_i$ by scaling it follows that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k) = Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ for each $k = 1, 2, \ldots, m$.

We state what we have shown as a theorem which is typically referred to as the **Gram-Schmidt Theorem**. When one uses the the result to obtain an **orthonormal basis** of a **subspace** of $\mathbb{R}^n$ we say that we are applying the **Gram-Schmidt Process**.

Theorem 8.2.1. *Let W be a* **subspace** *of $\mathbb{R}^n$ with* **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$. *Set* $\boldsymbol{x}_1 = \boldsymbol{w}_1$.

Assume that $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ have been defined with $k < m$. Set

$$\boldsymbol{x}_{k+1} = \boldsymbol{w}_{k+1} - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1 - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2}\boldsymbol{x}_2 - \cdots - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_k}{\boldsymbol{x}_k \cdot \boldsymbol{x}_k}\boldsymbol{x}_k$$

$$\boldsymbol{v}_{k+1} = \frac{1}{\parallel \boldsymbol{x}_{k+1} \parallel}\boldsymbol{x}_{k+1}.$$

Then the following hold:

1. The sequence of vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m)$ is an **orthonormal basis** *of W; and*

2. $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ for each $k = 1, 2, \ldots .m$.

Since every **subspace** of $\mathbb{R}^n$ has a **basis** we have the following theorem which is a consequence of the **Gram-Schmidt Theorem** (8.2.1).

Theorem 8.2.2. *Let W be a* **subspace** *of $\mathbb{R}^n$. Then W has an* **orthonormal basis**.

In section (5.5) we made use of the **Rank and Nullity Theorem** to show that if W is a subspace of $\mathbb{R}^n$ and $W^\perp$ is the **orthogonal complement** to W then $dim(W) + dim(W^\perp) = n$. We now make use of the Gram-Schmidt Theorem to give another proof:

Theorem 8.2.3. *Let W be a* **subspace** *of $\mathbb{R}^n$ and $W^\perp$ the* **orthogonal complement** *to W. Then $dim(W) + dim(W^\perp) = n$.*

Proof. Let $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ be a **basis** for W. Extend this to a **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ of $\mathbb{R}^n$. Apply the **Gram-Schmidt process** to $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ to obtain an **orthonormal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ for $\mathbb{R}^n$. By part 2) of **Theorem** (8.2.1), $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k) = W$.

Set $W' = Span(\boldsymbol{v}_{k+1}, \boldsymbol{v}_{k+2}, \dots, \boldsymbol{v}_n)$. Since $(\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n)$ is **linearly independent**, $(\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n)$ is a **basis** for W' and $dim(W') = n - k$.

For each pair (i, j) with $1 \le i \le k$ and $k + 1 \le j \le n$ we have $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$. From **Theorem** (2.6.7) we conclude that $W' \subset W^\perp$. We claim that $W' = W^\perp$ and the theorem will be a consequence of this. We therefore need to prove that $W^\perp \subset W'$.

Suppose that $\boldsymbol{x} \in W^\perp$. Since $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ is a **basis** for $\mathbb{R}^n$ we can write $\boldsymbol{x} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \dots c_n\boldsymbol{v}_n$ for scalars $c_1, c_2, \dots, c_n$. Choose $i, 1 \le i \le k$. Since $\boldsymbol{x} \in W^\perp$ and $\boldsymbol{v}_i \in W$ we can conclude that $0 = \boldsymbol{x} \cdot \boldsymbol{v}_i = c_1(\boldsymbol{v}_1 \cdot \boldsymbol{v}_i) + c_2(\boldsymbol{v}_2 \cdot \boldsymbol{v}_i) + \dots + c_k(\boldsymbol{v}_k \cdot \boldsymbol{v}_i) + c_{k+1}(\boldsymbol{v}_{k+1} \cdot \boldsymbol{v}_i) + \dots + c_n(\boldsymbol{v}_n \cdot \boldsymbol{v}_i) = c_i$ since $\boldsymbol{v}_j \cdot \boldsymbol{v}_i = 0$ for $i \ne j$ and $\boldsymbol{v}_i \cdot \boldsymbol{v}_i = 1$. Thus, $c_1 = c_2 = \dots = c_k = 0$ and $\boldsymbol{x} \in W'$ as required. $\square$

The QR Factorization of Matrices

We complete this section with another algorithm, which takes an $n \times m$ matrix A with **linearly independent** columns (equivalently, $rank(A) = m$) and expresses it as a product QR where Q is an $n \times m$ **orthonormal matrix** and R is an **invertible** upper triangular matrix. This is called a QR factorization of A. It used in computer algorithms, in particular, to find the **eigenvalues** of a matrix.

Definition 8.7. Let A be an $n \times m$ matrix with **linearly independent** columns. By a $QR-$***factorization of*** A we mean that A is the product of matrices Q and R where Q is an $n \times m$ **orthonormal matrix** and R is an $m \times m$ **invertible upper triangular matrix**.

We now prove that every $n \times m$ martrix A with **linearly independent** columns has a $QR-$factorization.

Theorem 8.2.4. *Let A be an $n \times m$ matrix with* **linearly independent** *columns. Then there is an $n \times m$* **orthonormal matrix** *Q and an $m \times m$* **invertible upper triangular matrix** *R such that $A = QR$.*

Proof. Let $A = (\boldsymbol{w}_1\ \boldsymbol{w}_2\ \dots\ \boldsymbol{w}_m)$ be an $n \times m$ matrix with **linearly independent** columns. Set $W = col(A)$. Then $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_m)$ is a **basis** $\mathcal{B}$ for W. Apply the **Gram-Schmidt process** to this **basis** to obtain an **orthonormal basis** $\mathcal{O} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ for W and let $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_m)$, an **orthonormal matrix**.

Let $[\boldsymbol{w}_k]_{\mathcal{O}} = \begin{pmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{mk} \end{pmatrix}$ be the **coordinate vector** of $\boldsymbol{w}_k$ with respect to $\mathcal{O}$. This means that $\boldsymbol{w}_k = r_{1k}\boldsymbol{v}_1 + r_{2k}\boldsymbol{v}_2 + \cdots + r_{mk}\boldsymbol{v}_m$ and consequently $\boldsymbol{w}_k = Q[\boldsymbol{w}_k]_{\mathcal{O}}$.

Thus, if $R = P_{\mathcal{B}\to\mathcal{O}} = ([\boldsymbol{w}_1]_{\mathcal{O}}\ [\boldsymbol{w}_2]_{\mathcal{O}}\ \dots\ [\boldsymbol{w}_m]_{\mathcal{O}})$ then

$$QR = (Q[\boldsymbol{w}_1]_{\mathcal{O}}\ Q[\boldsymbol{w}_2]_{\mathcal{O}}\ \dots\ Q[\boldsymbol{w}_m]_{\mathcal{O}}) = (\boldsymbol{w}_1\ \boldsymbol{w}_2\ \dots\ \boldsymbol{w}_m) = A.$$

Since R is a **change of basis matrix** , R is **invertible**. So, it remains to show that R is **upper triangular**. Recall, for each k, $Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k) = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$. Therefore, in the expression $\boldsymbol{w}_k = r_{1k}\boldsymbol{v}_1 + r_{2k}\boldsymbol{v}_2 + \cdots + r_{mk}\boldsymbol{v}_m$ all the coefficients r_{jk} with $j > k$ are zero: $r_{k+1,k} = r_{k+2,k} = \cdots = r_{m,k} = 0$. This implies that R is **upper triangular**, as required.

□

What You Can Now Do

1. Given a **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ for a **subspace** W of $\mathbb{R}^n$ obtain an **orthonormal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)$ for W.

2. Given an $n \times m$ matrix A with **linearly independent** columns, express A as a product QR where Q is an $n \times m$ **orthonormal matrix** and R is an **invertible** $m \times m$ **upper triangular** matrix.

Method (How To Do It)

Method 8.2.1. Given a **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ for a **subspace** W of $\mathbb{R}^n$ obtain an **orthonormal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k)$ for W.

We first define recursively a sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ of mutually **orthogonal vectors** such that $Span(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ for each $k = 1, 2, \ldots, m$.

Then the $\boldsymbol{v}_i$ are defined by **normalizing** the $\boldsymbol{x}_i$:

$$\boldsymbol{v}_i = \frac{1}{\parallel \boldsymbol{x}_i \parallel}\boldsymbol{x}_i, i = 1, 2, \ldots, m$$

The $\boldsymbol{x}_i$ are defined as follows:

i. $\boldsymbol{x}_1 = \boldsymbol{w}_1$.

ii. Assuming that $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ have been defined with $k < m$ we set

$$\boldsymbol{x}_{k+1} = \boldsymbol{w}_{k+1} - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1 - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2}\boldsymbol{x}_2 - \cdots - \frac{\boldsymbol{w}_{k+1} \cdot \boldsymbol{x}_k}{\boldsymbol{x}_k \cdot \boldsymbol{x}_k}\boldsymbol{x}_k$$

Example 8.2.1. Let $\boldsymbol{w}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\boldsymbol{w}_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$. Find an **orthonormal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ for $\mathbb{R}^2$ such that $Span(\boldsymbol{v}_1) = Span(\boldsymbol{w}_1)$ and $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$.

We begin by setting $\boldsymbol{x}_1 = \boldsymbol{w}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and then define $\boldsymbol{x}_2$ in terms of $\boldsymbol{x}_1$ and $\boldsymbol{w}_2$:

$$\boldsymbol{x}_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \frac{\begin{pmatrix} 3 \\ 4 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \end{pmatrix}}{\begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \end{pmatrix}} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \frac{18}{13}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \begin{pmatrix} \frac{36}{23} \\ \frac{54}{13} \end{pmatrix} = \begin{pmatrix} \frac{3}{13} \\ -\frac{2}{13} \end{pmatrix}$$

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{13}}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{pmatrix}, \boldsymbol{v}_2 = \sqrt{13}\begin{pmatrix} \frac{3}{13} \\ -\frac{2}{13} \end{pmatrix} = \begin{pmatrix} \frac{3}{\sqrt{13}} \\ -\frac{2}{\sqrt{13}} \end{pmatrix}$$

Example 8.2.2. Let $\boldsymbol{w}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}$, $\boldsymbol{w}_2 = \begin{pmatrix}1\\2\\3\end{pmatrix}$, and $\boldsymbol{w}_3 = \begin{pmatrix}2\\1\\1\end{pmatrix}$. Then $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is a **basis** of $\mathbb{R}^3$. Find an **orthonormal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ for $\mathbb{R}^3$ such that

$$Span(\boldsymbol{v}_1) = Span(\boldsymbol{w}_1), Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$$

$$Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3).$$

We first set $\boldsymbol{x}_1 = \boldsymbol{w}_1 = \begin{pmatrix}1\\1\\1\end{pmatrix}$.

$$\boldsymbol{x}_2 = \begin{pmatrix}1\\2\\3\end{pmatrix} - \frac{\begin{pmatrix}1\\1\\1\end{pmatrix} \cdot \begin{pmatrix}1\\2\\3\end{pmatrix}}{\begin{pmatrix}1\\1\\1\end{pmatrix} \cdot \begin{pmatrix}1\\1\\1\end{pmatrix}} \begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\2\\3\end{pmatrix} - \frac{6}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}-1\\0\\1\end{pmatrix}$$

$$\boldsymbol{x}_3 = \begin{pmatrix}2\\1\\1\end{pmatrix} - \frac{\begin{pmatrix}2\\1\\1\end{pmatrix} \cdot \begin{pmatrix}1\\1\\1\end{pmatrix}}{\begin{pmatrix}1\\1\\1\end{pmatrix} \cdot \begin{pmatrix}1\\1\\1\end{pmatrix}} \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{\begin{pmatrix}2\\1\\1\end{pmatrix} \cdot \begin{pmatrix}-1\\0\\1\end{pmatrix}}{\begin{pmatrix}-1\\0\\1\end{pmatrix} \cdot \begin{pmatrix}-1\\0\\1\end{pmatrix}} \begin{pmatrix}-1\\0\\1\end{pmatrix} =$$

$$\begin{pmatrix}2\\1\\1\end{pmatrix} - \frac{4}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{-1}{2}\begin{pmatrix}-1\\0\\1\end{pmatrix} = \begin{pmatrix}\frac{1}{6}\\-\frac{1}{3}\\\frac{1}{6}\end{pmatrix}$$

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{pmatrix}, \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\0\\1\end{pmatrix} = \begin{pmatrix}-\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{pmatrix},$$

$$\boldsymbol{v}_3 = \sqrt{6}\begin{pmatrix}\frac{1}{6}\\-\frac{1}{3}\\\frac{1}{6}\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{6}}\\-\frac{2}{\sqrt{6}}\\\frac{1}{\sqrt{6}}\end{pmatrix}$$

The **Gram-Schmidt process** applies to any **basis** of any **subspace** W of $\mathbb{R}^n$ not just to a **basis** of $\mathbb{R}^n$. The next two examples illustrates this:

Example 8.2.3. Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 3 \\ 7 \\ -5 \end{pmatrix}$ and set $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$. Find an **orthonormal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ for W such that $Span(\boldsymbol{v}_1) = Span(\boldsymbol{w}_1)$ and $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$.

$$\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}$$

$$\boldsymbol{x}_2 = \begin{pmatrix} 3 \\ 7 \\ -5 \end{pmatrix} - \frac{\begin{pmatrix} 3 \\ 7 \\ -5 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}}{\begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}} \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ -5 \end{pmatrix} - \frac{27}{9} \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} =$$

$$\begin{pmatrix} 3 \\ 7 \\ -5 \end{pmatrix} - \begin{pmatrix} 3 \\ 6 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$$\boldsymbol{v}_1 = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \\ -\frac{2}{3} \end{pmatrix}, \boldsymbol{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

We do one more example.

Example 8.2.4. Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 3 \\ 3 \\ 1 \\ 1 \end{pmatrix}$, and $\boldsymbol{w}_3 = \begin{pmatrix} 6 \\ 0 \\ 2 \\ -4 \end{pmatrix}$. Set $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$. Find an **orthonormal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ for W such that

$$Span(\boldsymbol{v}_1) = Span(\boldsymbol{w}_1), Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$$

$$Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3).$$

We begin by setting $\boldsymbol{x}_1 = \boldsymbol{w}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$.

$$\boldsymbol{x}_2 = \begin{pmatrix}3\\3\\1\\1\end{pmatrix} - \frac{\begin{pmatrix}3\\3\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}}{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \begin{pmatrix}3\\3\\1\\1\end{pmatrix} - \frac{8}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} =$$

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

$$\boldsymbol{x}_3 = \begin{pmatrix}6\\0\\2\\-4\end{pmatrix} - \frac{\begin{pmatrix}6\\0\\2\\-4\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}}{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{\begin{pmatrix}6\\0\\2\\-4\end{pmatrix}\cdot\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}}{\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} =$$

$$\begin{pmatrix}6\\0\\2\\-4\end{pmatrix} - \frac{4}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{8}{4}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = \begin{pmatrix}6\\0\\2\\-4\end{pmatrix} - \begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \begin{pmatrix}2\\2\\-2\\-2\end{pmatrix} = \begin{pmatrix}3\\-3\\3\\-3\end{pmatrix}.$$

$$\boldsymbol{v}_1 = \frac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{pmatrix}$$

$$\boldsymbol{v}_2 = \frac{1}{2}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\-\frac{1}{2}\\-\frac{1}{2}\end{pmatrix}$$

$$\boldsymbol{v}_3 = \frac{1}{5}\begin{pmatrix} 3 \\ -3 \\ 3 \\ -3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}.$$

Method 8.2.2. Given an $n \times m$ matrix A with **linearly independent** columns express A as a product QR where Q is an $n \times m$ **orthonormal matrix** and R is an **invertible** $m \times m$ **upper triangular** matrix.

If $A = (\boldsymbol{w}_1\ \boldsymbol{w}_2\ \dots\ \boldsymbol{w}_m)$ has **linearly independent** columns apply the Gram-Schmidt process, **Method** (8.2.1) to the **basis** $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_m)$ of $col(A)$ to obtain an **orthonormal basis** $\mathcal{O} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$ for $col(A)$ such that $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ for each $k = 1, 2, \dots, m$. Set $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \dots\ \boldsymbol{v}_m)$ and $R = P_{\mathcal{B}\to\mathcal{O}}$.

Remark 8.1. Since $\mathcal{O}$ is an **orthonormal basis** for $col(A)$ the **coordinate vector** for $\boldsymbol{u} \in col(A)$ with respect to $\mathcal{O}$ is $[\boldsymbol{u}]_{\mathcal{O}} = \begin{pmatrix} \boldsymbol{u} \cdot \boldsymbol{v}_1 \\ \boldsymbol{u} \cdot \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{u} \cdot \boldsymbol{v}_m \end{pmatrix}$.

In particular, $[\boldsymbol{w}_k]_{\mathcal{O}} = \begin{pmatrix} \boldsymbol{w}_k \cdot \boldsymbol{v}_1 \\ \boldsymbol{w}_k \cdot \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{w}_k \cdot \boldsymbol{v}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{w}_k \cdot \boldsymbol{v}_1 \\ \boldsymbol{w}_k \cdot \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{w}_k \cdot \boldsymbol{v}_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ since $\boldsymbol{w}_k \cdot \boldsymbol{v}_i = 0$ for $i > k$.

Example 8.2.5. Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. Obtain a **QR-factorization** for A.

Set $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}\right)$. Applying the Gram-Schmidt Process, **Method** (8.2.1), we obtain the **orthonormal basis** $\mathcal{O} = \left(\begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix}\right)$.

$$Q = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$[\begin{pmatrix} 1 \\ 2 \end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix} \sqrt{5} \\ 0 \end{pmatrix}, [\begin{pmatrix} 2 \\ 3 \end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix} \frac{8}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$$

$$R = \begin{pmatrix} \sqrt{5} & \frac{8}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix}$$

Example 8.2.6. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 8 \end{pmatrix}$. Obtain a **QR factorization** for A.

Set $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 8 \end{pmatrix} \right)$. Applying the Gram-Schmidt process, **Method** (8.2.1), to $\mathcal{B}$ we obtain the **orthonormal basis** $\mathcal{O} = \left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} -\frac{3}{\sqrt{26}} \\ -\frac{1}{\sqrt{26}} \\ \frac{4}{\sqrt{26}} \end{pmatrix} \right)$ for $col(A)$.

$$Q = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{3}{\sqrt{26}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{26}} \\ \frac{1}{\sqrt{3}} & \frac{4}{\sqrt{26}} \end{pmatrix}$$

$$[\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix} \sqrt{3} \\ 0 \end{pmatrix}, [\begin{pmatrix} 1 \\ 3 \\ 8 \end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix} \frac{12}{\sqrt{3}} \\ \frac{26}{\sqrt{26}} \end{pmatrix}$$

$$R = \begin{pmatrix} \sqrt{3} & \frac{12}{\sqrt{3}} \\ 0 & \sqrt{26} \end{pmatrix}$$

Example 8.2.7. Let $A = \begin{pmatrix} 1 & 3 & 3 \\ 1 & 3 & 3 \\ 2 & 0 & 9 \\ -2 & -2 & -3 \end{pmatrix}$. Obtain a **QR-factorization** of A.

Let $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 2 \\ -2 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 9 \\ -3 \end{pmatrix} \right)$ Applying the Gram-Schmidt process, **Method** (8.2.1), to $\mathcal{B}$ we obtain the **orthonormal basis** $\mathcal{O} = \left(\begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \\ \frac{2}{\sqrt{10}} \\ -\frac{2}{\sqrt{10}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{15}} \\ \frac{1}{\sqrt{15}} \\ \frac{2}{\sqrt{15}} \\ \frac{3}{\sqrt{15}} \end{pmatrix} \right)$ for $col(A)$.

$$Q = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{15}} \\ \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{15}} \\ \frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{15}} \\ -\frac{2}{\sqrt{10}} & 0 & \frac{3}{\sqrt{15}} \end{pmatrix}$$

$$[\begin{pmatrix} 1 \\ 1 \\ 2 \\ -2 \end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix} \sqrt{10} \\ 0 \\ 0 \end{pmatrix}, [\begin{pmatrix} 3 \\ 3 \\ 0 \\ -2 \end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix} \sqrt{10} \\ 2\sqrt{3} \\ 0 \end{pmatrix}, [\begin{pmatrix} 3 \\ 3 \\ 9 \\ -3 \end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix} 3\sqrt{10} \\ -\sqrt{3} \\ \sqrt{15} \end{pmatrix}$$

$$R = \begin{pmatrix} \sqrt{10} & \sqrt{10} & 3\sqrt{10} \\ 0 & 2\sqrt{3} & -\sqrt{3} \\ 0 & 0 & \sqrt{15} \end{pmatrix}$$

Example 8.2.8. Let $A = \begin{pmatrix} 1 & 3 & 6 \\ 1 & 3 & 0 \\ 1 & 1 & 2 \\ 1 & 1 & -4 \end{pmatrix}$. Obtain a **QR-factorization** of A.

Let $\mathcal{B} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 \\ 0 \\ 2 \\ -4 \end{pmatrix} \right)$. We encountered this sequence in **Example** (8.2.4).

The **orthonormal basis** obtained is $\mathcal{O} = \left(\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \right)$.

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

$$[\begin{pmatrix}1\\1\\1\\1\end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix}2\\0\\0\end{pmatrix}, [\begin{pmatrix}3\\3\\1\\1\end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix}4\\2\\0\end{pmatrix}, [\begin{pmatrix}6\\0\\2\\-4\end{pmatrix}]_{\mathcal{O}} = \begin{pmatrix}2\\4\\6\end{pmatrix}$$

$$R = \begin{pmatrix}2 & 4 & 2\\0 & 2 & 4\\0 & 0 & 6\end{pmatrix}$$

Exercises

In exercises 1-8 the given sequence is **linearly independent** and therefore a **basis** for the **subspace** W which it **spans**. Apply the Gram-Schmidt process, **Method** (8.2.1), to obtain an **orthonormal basis** for W.

1. $\left(\begin{pmatrix}1\\-2\\2\end{pmatrix}, \begin{pmatrix}5\\-3\\8\end{pmatrix}\right)$

2. $\left(\begin{pmatrix}1\\0\\3\end{pmatrix}, \begin{pmatrix}2\\3\\6\end{pmatrix}\right)$

3. $\left(\begin{pmatrix}1\\-1\\1\end{pmatrix}, \begin{pmatrix}5\\-6\\1\end{pmatrix}\right)$

4. $\left(\begin{pmatrix}2\\3\\-6\end{pmatrix}, \begin{pmatrix}10\\8\\-9\end{pmatrix}\right)$

5. $\left(\begin{pmatrix}1\\1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\5\\1\\-5\end{pmatrix}, \begin{pmatrix}9\\-3\\1\\-5\end{pmatrix}\right)$

6. $\left(\begin{pmatrix}1\\2\\1\\-2\end{pmatrix}, \begin{pmatrix}5\\2\\9\\-6\end{pmatrix}, \begin{pmatrix}-5\\6\\1\\-6\end{pmatrix}\right)$

7. $\left(\begin{pmatrix}1\\1\\2\\0\end{pmatrix}, \begin{pmatrix}1\\-3\\-2\\4\end{pmatrix}, \begin{pmatrix}2\\2\\1\\-3\end{pmatrix}\right)$

8. $\left(\begin{pmatrix}1\\2\\3\\0\end{pmatrix}, \begin{pmatrix}5\\7\\3\\-3\end{pmatrix}, \begin{pmatrix}1\\-1\\5\\7\end{pmatrix}\right)$.

In exercises 9 - 12 the matrix A has a factorization as QR for the given **orthonormal matrix** Q. Find the matrix R.

9. $A = \begin{pmatrix}1 & -1\\2 & 1\\2 & -5\end{pmatrix}, Q = \begin{pmatrix}\frac{1}{3} & 0\\\frac{2}{3} & \frac{1}{\sqrt{2}}\\\frac{2}{3} & -\frac{1}{\sqrt{2}}\end{pmatrix}$

10. $A = \begin{pmatrix}1 & 3 & 1\\1 & 3 & 5\\1 & 3 & -15\\-3 & -5 & 1\end{pmatrix}, Q = \begin{pmatrix}\frac{1}{\sqrt{12}} & \frac{1}{2} & \frac{1}{\sqrt{14}}\\\frac{1}{\sqrt{12}} & \frac{1}{2} & \frac{2}{\sqrt{14}}\\\frac{1}{\sqrt{12}} & \frac{1}{2} & -\frac{3}{\sqrt{14}}\\-\frac{3}{\sqrt{12}} & \frac{1}{2} & 0\end{pmatrix}$

11. $A = \begin{pmatrix}1 & 3 & 1\\1 & -5 & 1\\2 & 0 & -3\\2 & -4 & 7\end{pmatrix}, Q = \begin{pmatrix}\frac{1}{\sqrt{10}} & \frac{2}{\sqrt{10}} & \frac{1}{\sqrt{10}}\\\frac{1}{\sqrt{10}} & -\frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}}\\\frac{2}{\sqrt{10}} & \frac{1}{\sqrt{10}} & -\frac{2}{\sqrt{10}}\\\frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & \frac{2}{\sqrt{10}}\end{pmatrix}$

12. $A = \begin{pmatrix}1 & 2 & -1\\1 & -8 & 1\\1 & -3 & -6\\0 & 5 & 5\end{pmatrix}, Q = \begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0\\\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\\frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}}\\0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{pmatrix}$

In 13 - 18 find a **QR-factorization** for the given matrix A. See **Method** (8.2.2).

13. $A = \begin{pmatrix}1 & 6\\2 & 7\\2 & 8\end{pmatrix}$

14. $A = \begin{pmatrix}1 & 4\\1 & 0\\2 & 7\end{pmatrix}$

15. $A = \begin{pmatrix}2 & 7\\3 & 8\\6 & 10\end{pmatrix}$

16. $A = \begin{pmatrix} 1 & 3 & 1 \\ 0 & 2 & 1 \\ -1 & -5 & 4 \\ -1 & -1 & 0 \end{pmatrix}$

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

18. $A = \begin{pmatrix} 1 & 1 & 6 \\ 1 & 6 & 6 \\ 5 & 4 & 3 \\ 0 & 1 & 0 \end{pmatrix}$

In exercises 19 - 23 answer true or false and give an explanation.

19. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is an **orthonormal basis** of $\mathbb{R}^4$ and some of the vectors are multiplied by non-zero scalars then the resulting sequence is an **orthonormal basis** of $\mathbb{R}^4$.

20. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ is an **orthogonal basis** of $\mathbb{R}^4$ and each of the vectors in sequence **normalized** then the resulting sequence is an **orthonormal basis** of $\mathbb{R}^4$.

21. An **orthogonal sequence** of vectors in $\mathbb{R}^n$ is a **spanning sequence**.

22. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is an **orthogonal sequence** in $\mathbb{R}^3$ then it is a **basis** .

23. If $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is an **orthogonal sequence** of vectors in $\mathbb{R}^4$ then it can be extended to an **orthogonal basis** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4)$ for $\mathbb{R}^4$.

Challenge Exercises (Problems)

1. Assume $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly independent**. Set $\boldsymbol{x}_1 = \boldsymbol{w}_1, \boldsymbol{x}_2 = \boldsymbol{w}_2 - \frac{\boldsymbol{w}_2 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1$ and $\boldsymbol{x}_3 = \boldsymbol{w}_3 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{x}_1}{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}\boldsymbol{x}_1 - \frac{\boldsymbol{w}_3 \cdot \boldsymbol{x}_2}{\boldsymbol{x}_2 \cdot \boldsymbol{x}_2}\boldsymbol{x}_2$. Prove that $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ is an orthogonal sequence.

2. Let A be an $n \times m$ matrix with $rank(A) = m$, that is, the columns of A are **linearly independent**. Let $QR = A$ be the **QR-factorization** of A as obtained in **Theorem** (8.2.4). Prove that the **diagonal entries** of R are positive.

Here is an outline of the proof: Let $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ be the sequence of columns of A. Let $\mathcal{O}' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ be the **orthogonal sequence** first obtained in **Theorem** (8.2.4) and $\mathcal{O} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m)$ the **orthonormal sequence** obtained by **normalizing** each $\boldsymbol{x}_i$. Set $Q' = (\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \ldots \ \boldsymbol{x}_m)$ and $R' = P_{\mathcal{B} \to \mathcal{O}'}$. Show that $A = Q'R'$ and that the diagonal entries of R' are all 1. Then show that $Q' = QD$ where $D = diag(\frac{1}{\|\boldsymbol{x}_1\|}, \frac{1}{\|\boldsymbol{x}_2\|}, \ldots, \frac{1}{\|\boldsymbol{x}_m\|})$.)

3. Let $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$ be a two dimensional **subspace** of $\mathbb{R}^n$. Prove that there are precisely four **orthonormal bases** $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ for W such that $Span(\boldsymbol{v}_1) = Span(\boldsymbol{w}_1)$, $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$.

4. Let $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ be a three dimensional **subspace** of $\mathbb{R}^n$. Determine exactly how many **orthonormal bases** $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ there are for W such that $Span(\boldsymbol{v}_1) = Span(\boldsymbol{w}_1)$, $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$ and $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$.

Quiz Solutions

1. $\begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}$ Not right, see **Method** (2.6.1).

2. $\left(\begin{pmatrix} -4 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right)$.

3. $\boldsymbol{x} = \begin{pmatrix} \frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} -\frac{3}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{3}{2} \end{pmatrix}$. Not right, see **Method** (2.6.3).

4. $\boldsymbol{x}_1 = \begin{pmatrix} \frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$.

Not right, see the **discussion of orthogonal projection**.

8.3. Orthogonal Complements and Projections

In this section we introduce the make use of the **orthogonal complement** to a **subspace** of $\mathbb{R}^n$ to define the concept of an orthogonal projection. In turn, this is used to prove the **Best Approximation Theorem**.

Am I Ready for This Material

The following are important to an understanding of this section

orthogonal vectors

length (norm, magnitude) of a vector

orthogonal sequence

orthogonal basis for a subspace of $\mathbb{R}^n$

orthonormal sequence

orthonormal basis for a subspace of $\mathbb{R}^n$

direct sum decomposition

projection map $Proj_{(X,Y)}$

symmetric matrix

orthogonal complement of a subspace

projection of a vector onto the span of a given vector

projection of a vector orthogonal to a given vector

affine subspace of $\mathbb{R}^n$

Quiz

Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$, $\boldsymbol{w}_2 = \begin{pmatrix} 3 \\ 1 \\ -1 \\ -3 \end{pmatrix}$ and $\boldsymbol{w}_3 = \begin{pmatrix} 1 \\ 1 \\ -3 \\ 1 \end{pmatrix}$ and set $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$.

1. Find an **orthonormal basis** $\mathcal{O} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ for W such that $Span(\boldsymbol{v}_1) = Span(\boldsymbol{w}_1)$, $Span(\boldsymbol{v}_1, \boldsymbol{v}_2) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$ and $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$.

2. Find a **unit vector** in $W^{\perp}$.

3. If $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ compute the **change of basis matrix** $P_{\mathcal{B} \to \mathcal{O}}$.

4. If $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$ compute $\boldsymbol{u} - Proj_{\boldsymbol{v}_1}(\boldsymbol{u}) - Proj_{\boldsymbol{v}_2}(\boldsymbol{u}) - Proj_{\boldsymbol{v}_3}(\boldsymbol{u})$.

Quiz Solutions

New Concepts

In this section we introduce just a couple of new concepts:

orthogonal projection onto a subspace

distance of a vector to a subspace

Theory (Why It Works)

In section (2.6) we considered the following situation: we are given a non-zero vector $\boldsymbol{v} \in \mathbb{R}^n$ and a vector $\boldsymbol{u} \in \mathbb{R}^n$ and we wished to decompose $\boldsymbol{u}$ into a sum $\boldsymbol{x} + \boldsymbol{y}$ where $\boldsymbol{x}$ is a vector with the same direction as $\boldsymbol{v}$ ($\boldsymbol{x}$ is a multiple of $\boldsymbol{v}$) and $\boldsymbol{y}$ is a vector **orthogonal** to $\boldsymbol{v}$. We called the vector $\boldsymbol{x}$ in the direction of $\boldsymbol{v}$ the **orthogonal projection** of $\boldsymbol{u}$ onto $\boldsymbol{v}$. We denoted this by $Proj_{\boldsymbol{v}}(\boldsymbol{u})$ and demonstrated that $\boldsymbol{x} = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$.

We referred to the vector $\boldsymbol{y}$ as the **projection of u orthogonal to v** and denoted this by $Proj_{\boldsymbol{v}^\perp}(\boldsymbol{u})$. Of course, since $\boldsymbol{x}+\boldsymbol{y} = \boldsymbol{u}$ we must have $\boldsymbol{y} = \boldsymbol{u} - Proj_{\boldsymbol{v}}(\boldsymbol{u}) = \boldsymbol{u} - \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}$. This type of decomposition can be generalized to an arbitrary **subspace** W of $\mathbb{R}^n$.

Example 8.3.1. Let $W = Span\left(\begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\1\end{pmatrix}\right)$. Let $\boldsymbol{v} = \begin{pmatrix}v_1\\v_2\\v_3\\v_4\end{pmatrix}$ be an arbitrary vector in $\mathbb{R}^4$ and $\boldsymbol{w} = \begin{pmatrix}a\\a\\b\\b\end{pmatrix}$ a typical vector in W.

We want to find an a and b such that $\boldsymbol{v} - \boldsymbol{w}$ is **orthogonal** to W. Then $\boldsymbol{v} - \boldsymbol{w} = \begin{pmatrix}v_1 - a\\v_2 - a\\v_3 - b\\v_4 - b\end{pmatrix}$.

$$(\boldsymbol{v} - \boldsymbol{w}) \cdot \begin{pmatrix}1\\1\\0\\0\end{pmatrix} = \begin{pmatrix}v_1 - a\\v_2 - a\\v_3 - b\\v_4 - b\end{pmatrix} \cdot \begin{pmatrix}1\\1\\0\\0\end{pmatrix} =$$

$$(\boldsymbol{v} - \boldsymbol{w}) \cdot \begin{pmatrix}1\\1\\0\\0\end{pmatrix} = (v_1 - a) + (v_2 - a) + 0 + 0 = (v_1 + v_2) - 2a \tag{8.18}$$

$$(\boldsymbol{v}-\boldsymbol{w})\cdot\begin{pmatrix}0\\0\\1\\1\end{pmatrix}=\begin{pmatrix}v_1-a\\v_2-a\\v_3-b\\v_4-b\end{pmatrix}\cdot\begin{pmatrix}0\\0\\1\\1\end{pmatrix}$$

$$=0+0+(v_3-a)+(v_4-a)=(v_3+v_4)-2a \tag{8.19}$$

If $(\boldsymbol{v}-\boldsymbol{w})\perp W$ then from (8.18) and (8.19) we must have $a=\frac{v_1+v_2}{2}, b=\frac{v_3+v_4}{2}$. This gives the decomposition

$$\boldsymbol{v}=\begin{pmatrix}v_1\\v_2\\v_3\\v_4\end{pmatrix}=\begin{pmatrix}\frac{v_1+v_2}{2}\\\frac{v_1+v_2}{2}\\\frac{v_3+v_4}{2}\\\frac{v_3+v_4}{2}\end{pmatrix}+\begin{pmatrix}\frac{v_1-v_2}{2}\\\frac{-v_1+v_2}{2}\\\frac{v_3-v_4}{2}\\\frac{-v_3+v_4}{2}\end{pmatrix}.$$

Let W be a **subspace** of $\mathbb{R}^n$. Recall, in section (2.6) we introduced the concept of the **orthogonal complement** $W^\perp$ of $W: W^\perp=\{\boldsymbol{z}\in\mathbb{R}^n:\boldsymbol{w}\cdot\boldsymbol{z}=0 \text{ for all } \boldsymbol{w}\in W\}$. We collect some facts about $W^\perp$ that have already been demonstrated in previous sections (sometimes more than once).

Properties of $W^\perp$

Let W be a **subspace** of $\mathbb{R}^n$. Then the following hold:

1) $W^\perp$ is a **subspace** (see **Theorem** (2.6.7)).

2) $W\cap W^\perp=\{\mathbf{0}_n\}$. This follows immediately from the **positive definiteness** of the **dot product**.

3) $dim(W)+dim(W^\perp)=n$. This was shown in section (5.5) as **Theorem** (5.5.5) and again in section (8.2) as **Theorem** (8.2.3).

With these properties we can show that $W+W^\perp=\mathbb{R}^n$ and that every vector is uniquely a sum of a vector from W and a vector from $W^\perp$.

Theorem 8.3.1. *Let W be a* **subspace** *of $\mathbb{R}^n$. Then the following occur:*

1. $W+W^\perp=\mathbb{R}^n$, that is, for every vector $\boldsymbol{v}\in\mathbb{R}^n$ there is a vector $\boldsymbol{w}\in W$ and $\boldsymbol{z}\in W^\perp$ such that $\boldsymbol{v}=\boldsymbol{w}+\boldsymbol{z}$.

2. The decomposition of a vector $\boldsymbol{v}=\boldsymbol{w}+\boldsymbol{z}$ with $\boldsymbol{w}\in W,\boldsymbol{z}\in W^\perp$ is unique: If $\boldsymbol{w}_1,\boldsymbol{w}_2\in W,\boldsymbol{z}_1,\boldsymbol{z}_2\in W^\perp$ and $\boldsymbol{w}_1+\boldsymbol{z}_1=\boldsymbol{w}_2+\boldsymbol{z}_2$ then $\boldsymbol{w}_1=\boldsymbol{w}_2,\boldsymbol{z}_1=\boldsymbol{z}_2$.

Proof. Let $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ be a **basis** for W and $(\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_l)$ be a **basis** for $W^\perp$ (where $l = n-k$). We show that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k, \boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_l)$ is **linearly independent**. Suppose

$$a_1\boldsymbol{w}_1 + \dots a_k\boldsymbol{w}_k + b_1\boldsymbol{z}_1 + \dots + b_l\boldsymbol{z}_l = \mathbf{0}_n \tag{8.20}$$

Subtract $b_1\boldsymbol{z}_1 + \dots b_l\boldsymbol{z}_l$ from both sides of (8.20) to obtain

$$a_1\boldsymbol{w}_1 + \dots + a_k\boldsymbol{w}_k = -(b_1\boldsymbol{z}_1 + \dots + b_l\boldsymbol{z}_l) \tag{8.21}$$

The left hand side of (8.21) belongs to W and the right hand side to $W^\perp$. Since the two sides are equal, this vector belongs to $W \cap W^\perp$ which consists of only the zero vector, $\mathbf{0}_n$. Therefore

$$a_1\boldsymbol{w}_1 + \dots + a_k\boldsymbol{w}_k = b_1\boldsymbol{z}_1 + \dots + b_l\boldsymbol{z}_l = \mathbf{0}_n \tag{8.22}$$

Since $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ is a **basis** for W this sequence is **linearly independent** and, consequently, $a_1 = a_2 = \dots = a_k = 0$. Arguing in exactly the same way we also conclude that $b_1 = b_2 = \dots = b_l = 0$. So, $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_m, \boldsymbol{z}_1, \dots, \boldsymbol{z}_l)$ is **linearly independent** as claimed.

Since $k + l = n$ and $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_l)$ is **linearly independent** it follows that $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_l)$ is a **basis** for $\mathbb{R}^n$ by the **Theorem** (2.4.4). In particular, $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_l)$ **spans** $\mathbb{R}^n$.

Now if $\boldsymbol{v} \in \mathbb{R}^n$ then there exists scalars $a_1, \dots, a_k, b_1, \dots, b_l$ such that $\boldsymbol{v} = a_1\boldsymbol{w}_1 + \dots + a_k\boldsymbol{w}_k + b_1\boldsymbol{z}_1 + \dots + b_l\boldsymbol{z}_l$. Set $\boldsymbol{w} = a_1\boldsymbol{w}_1 + \dots + a_k\boldsymbol{w}_k$, an element of W and $\boldsymbol{z} = b_1\boldsymbol{z}_1 + \dots + b_l\boldsymbol{z}_l$, an element of $W^\perp$. Since $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{z}$ this gives the desired expression and we have proved part 1.

2) This is left as a **challenge exercise**. □

We make use of **Theorem** (8.3.1) in the next definition.

Definition 8.8. Let W be a **subspace** of $\mathbb{R}^n$ and let $\boldsymbol{v} \in \mathbb{R}^n$. Assume that $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{z}$ with $\boldsymbol{w} \in W, \boldsymbol{z} \in W^\perp$. Then $\boldsymbol{w}$ is called the ***orthogonal projection of $\boldsymbol{v}$ onto W*** and is denoted by $Proj_W(\boldsymbol{v})$. The vector $\boldsymbol{z}$ is called the ***projection of $\boldsymbol{v}$ orthogonal to W*** and is denoted by $Proj_{W^\perp}(\boldsymbol{v})$.

Remark 8.2. 1) Since $W \cap W^\perp = \{\mathbf{0}_n\}$ and $W + W^\perp = \mathbb{R}^n, \mathbb{R}^n = W \oplus W^\perp$, a **direct sum decomposition**. With such a decomposition we defined a **linear transformation** $Proj_{(W,W^\perp)}$. The **linear transformation** $Proj_{(W,W^\perp)}$ and $Proj_W$ are the same transformation as are $Proj_{(W^\perp,W)}$ and $Proj_{W^\perp}$.

2) For a vector $\boldsymbol{w} \in W, Proj_W(\boldsymbol{w}) = \boldsymbol{w}$. Since for any vector $\boldsymbol{v}, Proj_W(\boldsymbol{v}) \in W$ we conclude that $Proj_W^2(\boldsymbol{v}) = (Proj_W \circ Proj_W)(\boldsymbol{v}) = Proj_W(Proj_W(\boldsymbol{v})) = Proj_W(\boldsymbol{v})$.

In general if one is given a **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ for the **subspace** W and a vector $\boldsymbol{u} \in \mathbb{R}^n$ then it is fairly straightforward to compute $Proj_W(\boldsymbol{u})$. But first we need to show that a certain matrix is **invertible**.

Theorem 8.3.2. *A sequence $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ of vectors from $\mathbb{R}^n$ is* **linearly independent** *if and only if the matrix*

$$\begin{pmatrix} \boldsymbol{w}_1 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_1 & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_1 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_2 & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_2 \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_k & \boldsymbol{w}_2 \cdot \boldsymbol{w}_k & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_k \end{pmatrix}$$

is **invertible**.

Proof. All the elements of the general case are present when $k = 3$ so we deal with this situation.

Set $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ and $A = \begin{pmatrix} \boldsymbol{w}_1 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_3 \cdot \boldsymbol{w}_1 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_3 \cdot \boldsymbol{w}_2 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_3 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_3 & \boldsymbol{w}_3 \cdot \boldsymbol{w}_3 \end{pmatrix}$. Also, set $\boldsymbol{u} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3$.

We show that the vector $\boldsymbol{u} \in W^\perp$ if and only if $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \in null(A)$.

Now $\boldsymbol{u} \in W^\perp$ if and only if

$$\boldsymbol{u} \cdot \boldsymbol{w}_1 = \boldsymbol{u} \cdot \boldsymbol{w}_2 = \boldsymbol{u} \cdot \boldsymbol{w}_3 = 0.$$

Writing this out in terms of $\boldsymbol{u} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3$ we find that $\boldsymbol{u} \in W^\perp$ if and only if

$$(c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3) \cdot \boldsymbol{w}_1 = c_1(\boldsymbol{w}_1 \cdot \boldsymbol{w}_1) + c_2(\boldsymbol{w}_2 \cdot \boldsymbol{w}_1) + c_3(\boldsymbol{w}_3 \cdot \boldsymbol{w}_1) = 0 \quad (8.23)$$

$$(c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3)\cdot \boldsymbol{w}_2 = c_1(\boldsymbol{w}_1\cdot\boldsymbol{w}_2) + c_2(\boldsymbol{w}_2\cdot\boldsymbol{w}_2) + c_3(\boldsymbol{w}_3\cdot\boldsymbol{w}_2) = 0 \quad (8.24)$$

$$(c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3)\cdot \boldsymbol{w}_3 = c_1(\boldsymbol{w}_1\cdot\boldsymbol{w}_3) + c_2(\boldsymbol{w}_2\cdot\boldsymbol{w}_3) + c_3(\boldsymbol{w}_3\cdot\boldsymbol{w}_3) = 0 \quad (8.25)$$

Now, equations (8.23) - (8.25) can be written as

$$c_1\begin{pmatrix}\boldsymbol{w}_1\cdot\boldsymbol{w}_1\\ \boldsymbol{w}_1\cdot\boldsymbol{w}_2\\ \boldsymbol{w}_1\cdot\boldsymbol{w}_3\end{pmatrix} + c_2\begin{pmatrix}\boldsymbol{w}_2\cdot\boldsymbol{w}_1\\ \boldsymbol{w}_2\cdot\boldsymbol{w}_2\\ \boldsymbol{w}_2\cdot\boldsymbol{w}_3\end{pmatrix} + c_3\begin{pmatrix}\boldsymbol{w}_3\cdot\boldsymbol{w}_1\\ \boldsymbol{w}_3\cdot\boldsymbol{w}_2\\ \boldsymbol{w}_3\cdot\boldsymbol{w}_3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix} \quad (8.26)$$

In turn (8.26) can be written in matrix form

$$\begin{pmatrix}\boldsymbol{w}_1\cdot\boldsymbol{w}_1 & \boldsymbol{w}_2\cdot\boldsymbol{w}_1 & \boldsymbol{w}_3\cdot\boldsymbol{w}_1\\ \boldsymbol{w}_1\cdot\boldsymbol{w}_2 & \boldsymbol{w}_2\cdot\boldsymbol{w}_2 & \boldsymbol{w}_3\cdot\boldsymbol{w}_2\\ \boldsymbol{w}_1\cdot\boldsymbol{w}_3 & \boldsymbol{w}_2\cdot\boldsymbol{w}_3 & \boldsymbol{w}_3\cdot\boldsymbol{w}_3\end{pmatrix}\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix} \quad (8.27)$$

So, we have indeed shown that $\boldsymbol{u} \in W^\perp$ if and only if $\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix} \in null(A)$.

Assume now that A is **non-invertible**. By **Theorem** (3.4.15) it follows that $null(A) \neq \{\boldsymbol{0}_3\}$ and we can find a vector $\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix} \neq \{\boldsymbol{0}_3\}$ in $null(A)$. By what we have shown, the vector $\boldsymbol{u} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3$ belongs to $W^\perp$. Since $\boldsymbol{u} \in W$, this implies that $\boldsymbol{u}\cdot\boldsymbol{u} = 0$. By the **positive definiteness** of the **dot product** it must be the case that $\boldsymbol{u}$ is the zero vector, $\boldsymbol{u} = \boldsymbol{0}_3$. Since not all of c_1, c_2, c_3 are zero, this implies that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly dependent**.

Conversely, assume that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is **linearly dependent**. Then there are scalars, c_1, c_2, c_3, not all zero, such that $\boldsymbol{u} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3 = \boldsymbol{0}_3$. But then $\boldsymbol{u} \in W^\perp$ in which case $\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix} \in null(A)$ which implies that A is **non-invertible**. □

The next example shows how to find the **orthogonal projection** of a vector $\boldsymbol{u}$ onto a **subspace** W when given a **basis** of W.

Example 8.3.2. Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \\ 1 \end{pmatrix}$. Set $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ and $\boldsymbol{u} = \begin{pmatrix} 4 \\ 8 \\ -3 \\ 2 \end{pmatrix}$. Compute $Proj_W(\boldsymbol{u})$.

We want to find a vector $c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3$ such that $\boldsymbol{u} - (c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3)$ is in $W^\perp$. In particular, for each i we must have

$$[\boldsymbol{u} - (c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3)] \cdot \boldsymbol{w}_i =$$

$$\boldsymbol{u} \cdot \boldsymbol{w}_i - c_1(\boldsymbol{w}_1 \cdot \boldsymbol{w}_i) - c_2(\boldsymbol{w}_2 \cdot \boldsymbol{w}_i) - c_3(\boldsymbol{w}_3 \cdot \boldsymbol{w}_i) = 0 \tag{8.28}$$

Equation (8.28) is equivalent to

$$c_1(\boldsymbol{w}_1 \cdot \boldsymbol{w}_i) + c_2(\boldsymbol{w}_2 \cdot \boldsymbol{w}_i) + c_3(\boldsymbol{w}_3 \cdot \boldsymbol{w}_i) = \boldsymbol{u} \cdot \boldsymbol{w}_i \tag{8.29}$$

This means that $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$ is a **solution** to the **linear system** with **augmented matrix**

$$\left(\begin{array}{ccc|c} \boldsymbol{w}_1 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_3 \cdot \boldsymbol{w}_1 & \boldsymbol{u} \cdot \boldsymbol{w}_1 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_3 \cdot \boldsymbol{w}_2 & \boldsymbol{u} \cdot \boldsymbol{w}_2 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_3 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_3 & \boldsymbol{w}_3 \cdot \boldsymbol{w}_3 & \boldsymbol{u} \cdot \boldsymbol{w}_3 \end{array}\right) \tag{8.30}$$

It follows from **Theorem** (8.3.2) that this **linear system** has a unique **solution** which we now compute.

In our specific case we must solve the **linear system** with **augmented matrix**

$$\left(\begin{array}{ccc|c} 4 & 2 & 1 & 11 \\ 2 & 4 & -1 & 7 \\ 1 & -1 & 7 & 20 \end{array}\right) \tag{8.31}$$

This **linear system** has the unique **solution** $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. Thus, $Proj_W(\boldsymbol{u}) = \boldsymbol{w}_1 + 2\boldsymbol{w}_2 + 3\boldsymbol{w}_3 = \begin{pmatrix} 6 \\ -6 \\ 3 \\ 2 \end{pmatrix}$.

The example suggests the following theorem which provides a method for computing $Proj_W(\boldsymbol{u})$ when given a **basis** for the **subspace** W.

Theorem 8.3.3. *Let W be a **subspace** of $\mathbb{R}^n$ with **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k)$ and let $\boldsymbol{u}$ be a vector in $\mathbb{R}^n$. Then $Proj_W(\boldsymbol{u}) = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \ldots c_k\boldsymbol{w}_k$ where $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ is the unique **solution** to the **linear system** with **augmented matrix***

$$\left(\begin{array}{cccc|c} \boldsymbol{w}_1 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_1 & \ldots & \boldsymbol{w}_k \cdot \boldsymbol{w}_1 & \boldsymbol{u} \cdot \boldsymbol{w}_1 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_2 & \ldots & \boldsymbol{w}_k \cdot \boldsymbol{w}_2 & \boldsymbol{u} \cdot \boldsymbol{w}_2 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_k & \boldsymbol{w}_2 \cdot \boldsymbol{w}_k & \ldots & \boldsymbol{w}_k \cdot \boldsymbol{w}_k & \boldsymbol{u} \cdot \boldsymbol{w}_k \end{array}\right)$$

When given an **orthogonal basis** for W it is much easier to compute the **orthogonal projection** of a vector $\boldsymbol{v}$ onto W. We illustrate with an example before formulating this as a theorem.

Example 8.3.3. Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{w}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$ and $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$. Find the **orthogonal projection** of the vector $\boldsymbol{v} = \begin{pmatrix} 1 \\ 3 \\ -4 \\ 6 \end{pmatrix}$ onto W.

We claim that $Proj_W(\boldsymbol{v}) = \frac{\boldsymbol{v} \cdot \boldsymbol{w}_1}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\boldsymbol{w}_1 + \frac{\boldsymbol{v} \cdot \boldsymbol{w}_2}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\boldsymbol{w}_2$. We compute this vector

$$\frac{\boldsymbol{v} \cdot \boldsymbol{w}_1}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\boldsymbol{w}_1 + \frac{\boldsymbol{v} \cdot \boldsymbol{w}_2}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\boldsymbol{w}_2 = \frac{6}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \frac{2}{4}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \end{pmatrix} \tag{8.32}$$

The vector $\boldsymbol{w} = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \end{pmatrix}$ is a **linear combination** of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ and so in W. We need to show that the vector $\boldsymbol{v}-\boldsymbol{w}=\begin{pmatrix}-1\\1\\-5\\5\end{pmatrix}$ is **orthogonal** to $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$.

$$(\boldsymbol{v}-\boldsymbol{w})\cdot\boldsymbol{w}_1=\begin{pmatrix}-1\\1\\-5\\5\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}=-1+1-5+5=0 \tag{8.33}$$

$$(\boldsymbol{v}-\boldsymbol{w})\cdot\boldsymbol{w}_2=\begin{pmatrix}-1\\1\\-5\\5\end{pmatrix}\cdot\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}=-1+1+5-5=0 \tag{8.34}$$

Theorem 8.3.4. *Let W be a* **subspace** *of $\mathbb{R}^n$ and $(\boldsymbol{w}_1,\boldsymbol{w}_2,\dots,\boldsymbol{w}_k)$ be an* **orthogonal basis** *for W. Let $\boldsymbol{v}$ be a vector in $\mathbb{R}^n$. Then*

$$Proj_W(\boldsymbol{v})=\frac{\boldsymbol{v}\cdot\boldsymbol{w}_1}{\boldsymbol{w}_1\cdot\boldsymbol{w}_1}\boldsymbol{w}_1+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_2}{\boldsymbol{w}_2\cdot\boldsymbol{w}_2}\boldsymbol{w}_2+\cdots+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_k}{\boldsymbol{w}_k\cdot\boldsymbol{w}_k}\boldsymbol{w}_k$$

Proof. Set $\boldsymbol{w}=\frac{\boldsymbol{v}\cdot\boldsymbol{w}_1}{\boldsymbol{w}_1\cdot\boldsymbol{w}_1}\boldsymbol{w}_1+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_2}{\boldsymbol{w}_2\cdot\boldsymbol{w}_2}\boldsymbol{w}_2+\cdots+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_k}{\boldsymbol{w}_k\cdot\boldsymbol{w}_k}\boldsymbol{w}_k$, an element of W. We need to show that $\boldsymbol{v}-\boldsymbol{w}$ is perpendicular to $\boldsymbol{w}_1,\boldsymbol{w}_2,\dots,\boldsymbol{w}_k$.

We know from the **additive and scalar properties** of the **dot product** that $(\boldsymbol{v}-\boldsymbol{w})\cdot\boldsymbol{w}_i=\boldsymbol{v}\cdot\boldsymbol{w}_i-\boldsymbol{w}\cdot\boldsymbol{w}_i$ for each i. It also follows from the **additive and scalar properties** of the **dot product** that

$$\boldsymbol{w}\cdot\boldsymbol{w}_i=(\frac{\boldsymbol{v}\cdot\boldsymbol{w}_1}{\boldsymbol{w}_1\cdot\boldsymbol{w}_1}\boldsymbol{w}_1+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_2}{\boldsymbol{w}_2\cdot\boldsymbol{w}_2}\boldsymbol{w}_2+\cdots+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_k}{\boldsymbol{w}_k\cdot\boldsymbol{w}_k}\boldsymbol{w}_k)\cdot\boldsymbol{w}_i=$$

$$\frac{\boldsymbol{v}\cdot\boldsymbol{w}_1}{\boldsymbol{w}_1\cdot\boldsymbol{w}_1}(\boldsymbol{w}_1\cdot\boldsymbol{w}_i)+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_2}{\boldsymbol{w}_2\cdot\boldsymbol{w}_2}(\boldsymbol{w}_2\cdot\boldsymbol{w}_i)+\cdots+\frac{\boldsymbol{v}\cdot\boldsymbol{w}_k}{\boldsymbol{w}_k\cdot\boldsymbol{w}_k}(\boldsymbol{w}_k\cdot\boldsymbol{w}_i) \tag{8.35}$$

On the right hand side of (8.35) the only term which is non-zero is the term $\frac{\boldsymbol{w}\cdot\boldsymbol{w}_i}{\boldsymbol{w}_i\cdot\boldsymbol{w}_i}(\boldsymbol{w}_i\cdot\boldsymbol{w}_i)$ since for $j\neq i,\boldsymbol{w}_i\cdot\boldsymbol{w}_j=0$. Thus, $\boldsymbol{w}\cdot\boldsymbol{w}_i=\frac{\boldsymbol{v}\cdot\boldsymbol{w}_i}{\boldsymbol{w}_i\cdot\boldsymbol{w}_i}(\boldsymbol{w}_i\cdot\boldsymbol{w}_i)=\boldsymbol{v}\cdot\boldsymbol{w}_i$. Then

$$(\boldsymbol{v}-\boldsymbol{w})\cdot\boldsymbol{w}_i=\boldsymbol{v}\cdot\boldsymbol{w}_i-\boldsymbol{w}\cdot\boldsymbol{w}_i=\boldsymbol{v}\cdot\boldsymbol{w}_i-\boldsymbol{v}\cdot\boldsymbol{w}_i=0$$

as desired. □

You might recognize the expression $\frac{\boldsymbol{v}\cdot\boldsymbol{w}_i}{\boldsymbol{w}_i\cdot\boldsymbol{w}_i}\boldsymbol{w}_i$ as the **orthogonal projection** of the vector $\boldsymbol{v}$ onto $\boldsymbol{w}_i$. We therefore have the following:

Theorem 8.3.5. *Let* $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ *be an* **orthogonal basis** *for the* **subspace** W *of* $\mathbb{R}^n$ *and* $\boldsymbol{v}$ *a vector in* $\mathbb{R}^n$*. Then*

$$Proj_W(\boldsymbol{v}) = Proj_{\boldsymbol{w}_1}(\boldsymbol{v}) + Proj_{\boldsymbol{w}_2}(\boldsymbol{v}) + \dots + Proj_{\boldsymbol{w}_k}(\boldsymbol{v})$$

We next look at the matrix of an **orthogonal projection**.

Example 8.3.4. Let W be the **subspace** of **Example** (8.3.3). Find the **standard matrix** of $Proj_W$.

We find the **images** of the **standard basis vectors** making use of **Theorem** (8.3.4)

$$Proj_W(\boldsymbol{e}_1) = \frac{1}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \frac{1}{4}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\0\\0\end{pmatrix},$$

$$Proj_W(\boldsymbol{e}_2) = \frac{1}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \frac{1}{4}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\0\\0\end{pmatrix}$$

$$Proj_W(\boldsymbol{e}_3) = \frac{1}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{1}{4}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = \begin{pmatrix}0\\0\\\frac{1}{2}\\\frac{1}{2}\end{pmatrix},$$

$$Proj_W(\boldsymbol{e}_4) = \frac{1}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} - \frac{1}{4}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} = \begin{pmatrix}0\\0\\\frac{1}{2}\\\frac{1}{2}\end{pmatrix}$$

Therefore the **standard matrix** of $Proj_W$ is $A = \begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 0 & 0\\\frac{1}{2} & \frac{1}{2} & 0 & 0\\0 & 0 & \frac{1}{2} & \frac{1}{2}\\0 & 0 & \frac{1}{2} & \frac{1}{2}\end{pmatrix}$

Remark 8.3. We make two observations about the matrix of **Example** (8.3.4):

1) $A^2 = A$. This is to be expected since we know that $Proj_W^2 = Proj_W$ by **Remark** (8.2).

2) The matrix A is **symmetric**, $A^{Tr} = A$. We will show in general that the **standard matrix** of $Proj_W$ is always **symmetric**. This will be a consequence of the next theorem.

Theorem 8.3.6. *Let W be a* **subspace** *of $\mathbb{R}^n$ and assume that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ is an* **orthonormal basis** *of W. Let A be the $n \times k$ matrix whose columns are the vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k, A = (\boldsymbol{w}_1 \ \boldsymbol{w}_2 \ \dots \ \boldsymbol{w}_k)$. Then the* **standard matrix** *of $Proj_W$ is AA^{Tr}.*

Proof. The **standard matrix** of $Proj_W$ is the matrix with columns the **images** of the **standard basis vectors**, $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$:

$$(Proj_W(\boldsymbol{e}_1) \ Proj_W(\boldsymbol{e}_2) \ \dots \ Proj_W(\boldsymbol{e}_n))$$

We can use **Theorem** (8.3.5) to compute $Proj_W(\boldsymbol{e}_j)$, making use of the fact that each $\boldsymbol{w}_i$ is a **unit vector**:

$$Proj_W(\boldsymbol{e}_j) = Proj_{\boldsymbol{w}_1}(\boldsymbol{e}_j) + Proj_{\boldsymbol{w}_2}(\boldsymbol{e}_j) + \dots + Proj_{\boldsymbol{w}_k}(\boldsymbol{e}_j) =$$

$$(\boldsymbol{e}_j \cdot \boldsymbol{w}_1)\boldsymbol{w}_1 + (\boldsymbol{e}_j \cdot \boldsymbol{w}_2)\boldsymbol{w}_2 + \dots + (\boldsymbol{e}_j \cdot \boldsymbol{w}_k)\boldsymbol{w}_k \tag{8.36}$$

We deduce from (8.36) and the definition of the product of an $n \times k$ matrix and a $k-$vector that

$$Proj_W(\boldsymbol{e}_j) = (\boldsymbol{w}_1 \ \boldsymbol{w}_2 \ \dots \ \boldsymbol{w}_k)\begin{pmatrix} \boldsymbol{e}_j \cdot \boldsymbol{w}_1 \\ \boldsymbol{e}_j \cdot \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{e}_j \cdot \boldsymbol{w}_k \end{pmatrix} = A\begin{pmatrix} \boldsymbol{e}_j \cdot \boldsymbol{w}_1 \\ \boldsymbol{e}_j \cdot \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{e}_j \cdot \boldsymbol{w}_k \end{pmatrix} \tag{8.37}$$

Since $A\begin{pmatrix} \boldsymbol{e}_j \cdot \boldsymbol{w}_1 \\ \boldsymbol{e}_j \cdot \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{e}_j \cdot \boldsymbol{w}_k \end{pmatrix}$ is the j^{th} column of the **standard matrix** of $Proj_W$ we deduce that the **standard matrix** is $A \begin{pmatrix} \boldsymbol{e}_1 \cdot \boldsymbol{w}_1 & \boldsymbol{e}_2 \cdot \boldsymbol{w}_1 & \dots & \boldsymbol{e}_n \cdot \boldsymbol{w}_1 \\ \boldsymbol{e}_1 \cdot \boldsymbol{w}_2 & \boldsymbol{e}_2 \cdot \boldsymbol{w}_2 & \dots & \boldsymbol{e}_n \cdot \boldsymbol{w}_2 \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{e}_1 \cdot \boldsymbol{w}_k & \boldsymbol{e}_2 \cdot \boldsymbol{w}_k & \dots & \boldsymbol{e}_n \cdot \boldsymbol{w}_k \end{pmatrix} = AB.$

The i^{th} row of the matrix B is $\boldsymbol{w}_i^{Tr}$ and therefore $B = A^{Tr}$. Consequently, the **standard matrix** of $Proj_W$ is AA^{Tr} as claimed.

□

Remark 8.4. 1) It is an immediate corollary of **Theorem** (8.3.6) that the **standard matrix** of an **orthogonal projection** is **symmetric**.

2) The two properties $Q^2 = Q, Q = Q^{Tr}$ characterizes the matrix of an **orthogonal projection**. This is assigned as a **challenge exercise**.

Recall in section (2.6) we showed that for a non-zero vector $\boldsymbol{v}$ in $\mathbb{R}^n$ and any vector $\boldsymbol{u}$ we could interpret $Proj_{\boldsymbol{v}}(\boldsymbol{u})$ as the vector in $Span(\boldsymbol{v})$ which is "closest" to $\boldsymbol{u}$ in the sense that $\| \boldsymbol{u} - Proj_{\boldsymbol{v}}(\boldsymbol{u}) \|$ is strictly less than $\| \boldsymbol{u} - \boldsymbol{x} \|$ for any vector $\boldsymbol{x} \in Span(\boldsymbol{v}), \boldsymbol{x} \neq Proj_{\boldsymbol{v}}(\boldsymbol{u})$. This property is also true of $Proj_W(\boldsymbol{u})$ which is the subject of our next result which we refer to as the **Best Approximation Theorem**.

Theorem 8.3.7. *Let W be a* **subspace** *of $\mathbb{R}^n$ and $\boldsymbol{u}$ an $n-$vector. Then for any vector $\boldsymbol{w} \in W, \boldsymbol{w} \neq Proj_W(\boldsymbol{u})$ we have*

$$\| \boldsymbol{u} - Proj_W(\boldsymbol{u}) \| < \| \boldsymbol{u} - \boldsymbol{w} \|$$

Proof. Set $\hat{\boldsymbol{w}} = Proj_W(\boldsymbol{u})$. Then the vector $\boldsymbol{u} - \hat{\boldsymbol{w}} \in W^{\perp}$ and so **orthogonal** to every vector in W. In particular, $\boldsymbol{u} - \hat{\boldsymbol{w}}$ is **orthogonal** to $\hat{\boldsymbol{w}} - \boldsymbol{w}$.

Now $\boldsymbol{u} - \boldsymbol{w} = (\boldsymbol{u} - \hat{\boldsymbol{w}}) + (\hat{\boldsymbol{w}} - \boldsymbol{w})$. Since $\boldsymbol{u} - \hat{\boldsymbol{w}}$ is **orthogonal** to $\hat{\boldsymbol{w}} - \boldsymbol{w}$ we have

$$\| \boldsymbol{u} - \boldsymbol{w} \|^2 = \| (\boldsymbol{u} - \hat{\boldsymbol{w}}) + (\hat{\boldsymbol{w}} - \boldsymbol{w}) \|^2 = \| \boldsymbol{u} - \hat{\boldsymbol{w}} \|^2 + \| \hat{\boldsymbol{w}} - \boldsymbol{w} \|^2 \tag{8.38}$$

by the **Pythagorean theorem**. Since $\boldsymbol{w} \neq Proj_W(\boldsymbol{u}) = \hat{\boldsymbol{w}}, \hat{\boldsymbol{w}} - \boldsymbol{w} \neq 0$ and consequently, $\| \hat{\boldsymbol{w}} - \boldsymbol{w} \| \neq 0$. From (8.38) we conclude that

$$\| \boldsymbol{u} - \boldsymbol{w} \|^2 > \| \boldsymbol{u} - \hat{\boldsymbol{w}} \|^2 \tag{8.39}$$

from which the result immediately follows by taking square roots. □

Definition 8.9. Let W be a subspace of $\mathbb{R}^n$ and let $\boldsymbol{u} \in \mathbb{R}^n$. The ***distance of $\boldsymbol{u}$ to W*** is the minimum of $\{\parallel \boldsymbol{u} - \boldsymbol{w} \parallel : \boldsymbol{w} \in W\}$, that is, the shortest distance of the vector $\boldsymbol{u}$ to a vector in W. By Theorem (8.3.7) this is $\parallel \boldsymbol{u} - Proj_W(\boldsymbol{u}) \parallel$. We denote the distance of the vector $\boldsymbol{u}$ to the subspace W by $dist(\boldsymbol{u}, W)$.

More generally, the ***distance*** of a vector $\boldsymbol{u}$ to an affine subspace of $\mathbb{R}^n$ $W + \boldsymbol{p}$ is defined to be the minimum of $dist(\boldsymbol{u}, \boldsymbol{w} + \boldsymbol{p})$ where $\boldsymbol{w}$ is an arbitrary vector in W. This is denoted by $dist(\boldsymbol{u}, W + \boldsymbol{p})$.

Example 8.3.5. Let $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$ and set $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$. Find the distance of the vector $\boldsymbol{u} = \begin{pmatrix} 9 \\ 3 \\ 3 \\ 1 \end{pmatrix}$ to W.

The sequence $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an orthogonal basis of W. By normalizing we obtain an orthonormal basis of W :

$$\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$$

The standard matrix of $Proj_W$ is

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

Then

$$Proj_W(\boldsymbol{u}) = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{pmatrix} \begin{pmatrix} 9 \\ 3 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \\ 4 \\ 0 \end{pmatrix}$$

$$\boldsymbol{u} - Proj_W(\boldsymbol{u}) = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \parallel \boldsymbol{u} - Proj_W(\boldsymbol{u}) \parallel = 2$$

What You Can Now Do

1. Given a **subspace** W of $\mathbb{R}^n$ and an $n-$vector $\boldsymbol{u}$, compute $Proj_W(\boldsymbol{u})$, the **orthogonal projection** of $\boldsymbol{u}$ onto W.

2. Given a **subspace** W of $\mathbb{R}^n$ compute the **standard matrix** of $Proj_W$.

3. Given a **subspace** W of $\mathbb{R}^n$ and an $n-$vector $\boldsymbol{u}$, compute $dist(\boldsymbol{u}, W)$, the **distance** of $\boldsymbol{u}$ to W.

4. Given an **affine subspace** of $\mathbb{R}^n$ and an $n-$vector $\boldsymbol{u}$ compute the **distance** of $\boldsymbol{u}$ to $W + \boldsymbol{p}$.

Method (How To Do It)

Method 8.3.1. Given a **subspace** W of $\mathbb{R}^n$ and an $n-$vector $\boldsymbol{u}$, compute $Proj_W(\boldsymbol{u})$, the **orthogonal projection** of $\boldsymbol{u}$ onto W.

Presumably W is given as a **span** of a sequence of vectors (which may or may not be a **basis** for W), $W = Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$. If necessary, use **Method** (5.5.1) or **Method** (5.5.3) to obtain a basis $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ for W (it may be that $k = m$).

Thus, we may assume that we are given a **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ for W. Set

$$A = \begin{pmatrix} \boldsymbol{w}_1 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_1 & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_1 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_2 & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_2 \\ \vdots & \vdots & \dots & \vdots \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_k & \boldsymbol{w}_2 \cdot \boldsymbol{w}_k & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_k \end{pmatrix}$$

then $Proj_W(\boldsymbol{u}) = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \dots + c_k\boldsymbol{w}_k$ where $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$ is the unique **solution** to the **linear system** with **augmented matrix**

$$\left(\begin{array}{cccc|c} \boldsymbol{w}_1 \cdot \boldsymbol{w}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_1 & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_1 & \boldsymbol{u} \cdot \boldsymbol{w}_1 \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_2 & \boldsymbol{w}_2 \cdot \boldsymbol{w}_2 & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_2 & \boldsymbol{u} \cdot \boldsymbol{w}_2 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \boldsymbol{w}_1 \cdot \boldsymbol{w}_k & \boldsymbol{w}_2 \cdot \boldsymbol{w}_k & \dots & \boldsymbol{w}_k \cdot \boldsymbol{w}_k & \boldsymbol{u} \cdot \boldsymbol{w}_k \end{array}\right)$$

In the event that $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ is an **orthogonal sequence** the matrix A is **diagonal** with the entries $\boldsymbol{w}_i \cdot \boldsymbol{w}_i$ on the diagonal. In this case $Proj_W(\boldsymbol{u})$ is easy to compute:

$$Proj_W(\boldsymbol{u}) = \frac{\boldsymbol{u} \cdot \boldsymbol{w}_1}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\boldsymbol{w}_1 + \frac{\boldsymbol{u} \cdot \boldsymbol{w}_2}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\boldsymbol{w}_2 + \dots + \frac{\boldsymbol{u} \cdot \boldsymbol{w}_k}{\boldsymbol{w}_k \cdot \boldsymbol{w}_k}\boldsymbol{w}_k$$

In the even more advantageous situation where $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ is an **orthonormal basis** for W then the matrix A is the **identity matrix** and then

$$Proj_W(\boldsymbol{u}) = (\boldsymbol{u} \cdot \boldsymbol{w}_1)\boldsymbol{w}_1 + (\boldsymbol{u} \cdot \boldsymbol{w}_2)\boldsymbol{w}_2 + \dots + (\boldsymbol{u} \cdot \boldsymbol{w}_k)\boldsymbol{w}_k$$

Example 8.3.6. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix}$ and $W = Span(\boldsymbol{v}_1, \boldsymbol{v}_2)$. Find $Proj_W(\boldsymbol{u})$ for the vector $\boldsymbol{u} = \begin{pmatrix} 3 \\ 1 \\ -7 \end{pmatrix}$.

We need to solve the **linear system** with **augmented matrix**

$$\begin{pmatrix} 9 & 9 & | & -9 \\ 9 & 17 & | & -25 \end{pmatrix}$$

The unique **solution** to this **linear system** is $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$. Therefore

$$Proj_W(\boldsymbol{u}) = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - 2\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ -6 \end{pmatrix}$$

Example 8.3.7. Let $\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$ so that $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a **basis** for the space $W = \{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} : x_1 + x_2 + x_3 + x_4 = 0\}$. Let $\boldsymbol{u} = \begin{pmatrix} 1 \\ -4 \\ 3 \\ -4 \end{pmatrix}$. Compute $Proj_W(\boldsymbol{u})$.

We need to compute the unique **solution** to the **linear system** with **augmented matrix**

$$\begin{pmatrix} 4 & 2 & 2 & | & -2 \\ 2 & 2 & 1 & | & 5 \\ 2 & 1 & 2 & | & -2 \end{pmatrix}$$

The **solution** to this **linear system** is $\begin{pmatrix} -3 \\ 6 \\ -1 \end{pmatrix}$. Therefore

$$Proj_W(\boldsymbol{u}) = -3\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} + 6\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \\ 4 \\ -3 \end{pmatrix}$$

Example 8.3.8. Let W and $\boldsymbol{u}$ be as in **Example** (8.3.7) but we now use the **orthogonal basis** $(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ where $\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$ for W.

Now we can compute the projection directly:

$$Proj_W(\boldsymbol{u}) = \frac{\boldsymbol{u} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{u} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2}\boldsymbol{u}_2 + \frac{\boldsymbol{u} \cdot \boldsymbol{u}_3}{\boldsymbol{u}_3 \cdot \boldsymbol{u}_3}\boldsymbol{u}_3 =$$

$$\frac{-2}{4}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} + \frac{12}{4}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} + \frac{-2}{4}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \\ 4 \\ -3 \end{pmatrix}$$

Method 8.3.2. Given a **subspace** W of $\mathbb{R}^n$ compute the **standard matrix** of $Proj_W$.

By given W we mean that W is expressed as $Span(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m)$. These vectors may or may not be **linearly independent**. If necessary, use **Method** (5.5.1) or **Method** (5.5.3) to obtain a basis $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ for W (it may be that $k = m$). So we may assume we are given $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ where $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k)$ is **linearly independent**.

If necessary, apply the Gram-Schmidt process, **Method** (8.2.1), to the **basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k)$ to order to obtain an **orthonormal basis** $(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_k)$. Let A be the matrix $(\boldsymbol{x}_1\ \boldsymbol{x}_2\ \dots\ \boldsymbol{x}_k)$ and set $Q = AA^{Tr}$. Then Q is the **standard matrix** of $Proj_W$.

Example 8.3.9. Let $W = Span\left(\begin{pmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{pmatrix}, \begin{pmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\\ 0\end{pmatrix}\right)$. Compute the **standard matrix** of $Proj_W$.

Set $A = \begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & 0\end{pmatrix}$. Then the **standard matrix** of $Proj_W$ is

$$Q = AA^T = \begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & 0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\end{pmatrix} = \begin{pmatrix}\frac{5}{6} & -\frac{1}{6} & \frac{1}{3}\\ -\frac{1}{6} & \frac{5}{6} & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix}$$

Example 8.3.10. Let $W = Span\left(\begin{pmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ 0\end{pmatrix}, \begin{pmatrix}\frac{1}{\sqrt{3}}\\ -\frac{1}{\sqrt{3}}\\ 0\\ \frac{1}{\sqrt{3}}\end{pmatrix}, \begin{pmatrix}\frac{1}{\sqrt{3}}\\ 0\\ -\frac{1}{\sqrt{3}}\\ -\frac{1}{\sqrt{3}}\end{pmatrix}\right)$. Compute the **standard matrix** of W.

Set $A = \begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}}\\ 0 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}}\end{pmatrix}$. Then the **standard matrix** of $Proj_W$ is

$$Q = AA^T = \begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}}\\ 0 & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}} & 0 & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}}\end{pmatrix} =$$

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

Method 8.3.3. Given a **subspace** W of $\mathbb{R}^n$ and an $n-$vector $\boldsymbol{u}$, compute $dist(\boldsymbol{u}, W)$, the **distance** of $\boldsymbol{u}$ to W.

Compute $Proj_W(\boldsymbol{u})$. The **distance** of $\boldsymbol{u}$ to W is $\parallel \boldsymbol{u} - Proj_W(\boldsymbol{u}) \parallel$.

Example 8.3.11. Let $W = Span\left(\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}\right)$ and $\boldsymbol{u} = \begin{pmatrix} 10 \\ 2 \\ 4 \\ 0 \end{pmatrix}$. Find the **distance** of $\boldsymbol{u}$ to W.

The given sequence is an **orthogonal basis** of W. We **normalize** to obtain an **orthonormal basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2)$ where $\boldsymbol{w}_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$ and $\boldsymbol{w}_2 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$. We can now compute the **orthogonal projection** of $\boldsymbol{u}$ onto W directly:

$$Proj_W(\boldsymbol{u}) = (\boldsymbol{u} \cdot \boldsymbol{w}_1)\boldsymbol{w}_1 + (\boldsymbol{u} \cdot \boldsymbol{w}_2)\boldsymbol{w}_2 = \begin{pmatrix} 5 \\ -1 \\ 1 \\ -5 \end{pmatrix}$$

Then $\boldsymbol{u} - Proj_W(\boldsymbol{u}) = \begin{pmatrix} 10 \\ 2 \\ 4 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ -1 \\ 1 \\ -5 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \\ 3 \\ 5 \end{pmatrix}$. The **distance** of $\boldsymbol{u}$ to W is

$$\parallel \begin{pmatrix} 5 \\ 3 \\ 3 \\ 5 \end{pmatrix} \parallel = \sqrt{25 + 9 + 9 + 25} = \sqrt{68}.$$

Example 8.3.12. Now let $\boldsymbol{v}_3 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$ and set $X = Span(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{v}_3)$ where

$(\boldsymbol{w}_1, \boldsymbol{w}_2)$ is the **orthonormal basis** for the **subspace** W of **Example** (8.3.11). Compute the **distance** of $\begin{pmatrix} 10 \\ 2 \\ 4 \\ 0 \end{pmatrix}$ to X.

The vector $\boldsymbol{v}_3$ is **orthogonal** to $\boldsymbol{w}_1, \boldsymbol{w}_2$. If we set $\boldsymbol{w}_3 = \frac{1}{2}\boldsymbol{v}_3 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$ then $(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$ is an **orthonormal basis** of X. We do not have to start over to compute $Proj_X(\boldsymbol{u})$ since

$$Proj_X(\boldsymbol{u}) = Proj_W(\boldsymbol{u}) + Proj_{\boldsymbol{w}_3}(\boldsymbol{u}) = Proj_W(\boldsymbol{u}) + (\boldsymbol{u} \cdot \boldsymbol{w}_3)\boldsymbol{w}_3$$

We have already computed $Proj_W(\boldsymbol{u}) = \begin{pmatrix} 5 \\ -1 \\ 1 \\ -5 \end{pmatrix}$ in **Example** (8.3.11). It remains to compute $Proj_{\boldsymbol{w}_3}(\boldsymbol{u})$ and $Proj_X(\boldsymbol{u})$:

$$Proj_{\boldsymbol{w}_3}(\boldsymbol{u}) = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

and therefore

$$Proj_X(\boldsymbol{u}) = \begin{pmatrix} 5 \\ -1 \\ 1 \\ -5 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ -2 \\ 0 \\ -4 \end{pmatrix}$$

Thus, $\boldsymbol{u} - Proj_X(\boldsymbol{u}) = \begin{pmatrix} 10 \\ 2 \\ 4 \\ 0 \end{pmatrix} - \begin{pmatrix} 6 \\ -2 \\ 0 \\ -4 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 4 \\ 4 \end{pmatrix}$

Therefore the **distance** of $\boldsymbol{u}$ to X is $\| \begin{pmatrix} 4 \\ 4 \\ 4 \\ 4 \end{pmatrix} \| = 8$.

Method 8.3.4. Given an **affine subspace** of $\mathbb{R}^n$ and an $n-$vector $\boldsymbol{u}$ compute the distance of $\boldsymbol{u}$ to $W + \boldsymbol{p}$.

A typical vector in $W + \boldsymbol{p}$ has the form $\boldsymbol{w} + \boldsymbol{p}$. The **distance** of $\boldsymbol{u}$ to $\boldsymbol{w} + \boldsymbol{p}$ is $\| \boldsymbol{u} - (\boldsymbol{w} + \boldsymbol{p}) \| = \| (\boldsymbol{u} - \boldsymbol{p}) - \boldsymbol{w} \|$. Therefore the **distance** of $\boldsymbol{u}$ to the **affine subspace** $W + \boldsymbol{p}$ is the same as the distance of $\boldsymbol{u} - \boldsymbol{p}$ to W. Apply **Method** (8.3.3) to the space W and the vector $\boldsymbol{u} - \boldsymbol{p}$.

Example 8.3.13. Find the **distance** of the vector $\boldsymbol{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ from the plane

$$x + 2y + 2z = 4 \tag{8.40}$$

In the usual fashion we express the **solution set** of the equation $x + 2y + 2z = 4$ as an **affine subspace** of $\mathbb{R}^3$ using **Method** (3.2.3). The **solution space** to the **homogeneous linear equation**

$$x + 2y + 2z = 0$$

is $Span\left(\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}\right)$. A specific solution to Equation (8.40) is $\boldsymbol{p} = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}$ and the solution set of Equation (8.40) is $Span\left(\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}\right) + \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}$.

The distance of $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ from $Span\left(\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}\right) + \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}$ is the distance of $\boldsymbol{u} - \boldsymbol{p} = \begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix}$ from $W = Span\left(\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}\right)$. We have to find the projection of $\boldsymbol{u} - \boldsymbol{p}$ onto W.

The sequence $\left(\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}\right)$ is an **orthogonal basis** of W. Making use of this we find that

$$Proj_W(\boldsymbol{u} - \boldsymbol{p}) = \begin{pmatrix} -\frac{28}{9} \\ \frac{7}{9} \\ \frac{7}{9} \end{pmatrix}$$

Then $(\boldsymbol{u}-\boldsymbol{p}) - Proj_W(\boldsymbol{u}-\boldsymbol{p}) = \begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} -\frac{28}{9} \\ \frac{7}{9} \\ \frac{7}{9} \end{pmatrix} = \begin{pmatrix} \frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{pmatrix}$.

The distance of $\boldsymbol{u}$ from $W+\boldsymbol{p}$ is the length of $\begin{pmatrix} \frac{1}{9} \\ \frac{2}{9} \\ \frac{2}{9} \end{pmatrix}$ which is $\frac{1}{3}$.

Exercises

In exercises 1 - 6 compute the **orthogonal projection**, $Proj_W(\boldsymbol{u})$, for the given vector $\boldsymbol{u}$ and **subspace** W. Note the **spanning sequence** given for W may not be **linearly independent**.

1. $W = Span\left(\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 6 \\ 1 \\ -1 \end{pmatrix}$

2. $W = Span\left(\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix}$.

3. $W = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ 3 \\ -3 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 3 \\ 1 \\ 7 \\ -5 \end{pmatrix}$

4. $W = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ -3 \\ 3 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 5 \\ -3 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 3 \\ 1 \\ 7 \\ -5 \end{pmatrix}$

5. $W = Span\left(\begin{pmatrix} 1 \\ 2 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 8 \\ 1 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 7 \\ 5 \\ 5 \\ 4 \end{pmatrix}$

6. $W = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 3 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ 1 \\ -1 \end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix} 9 \\ -7 \\ 2 \\ 1 \end{pmatrix}$

In exercises 7 - 12, for the given subspace W, determine the **standard matrix** of the **orthogonal projection** onto W, $Proj_W$, for the given **subspace** W.

7. The space of exercise 1.

8. The space of exercise 2.

9. The space of exercise 3.

10. The space of exercise 4.

11. The space of exercise 5.

12. The space of exercise 6.

In exercises 13 - 20 find the **distance** of the given vector $\boldsymbol{u}$ to the **subspace** W.

13. $W = Span\left(\begin{pmatrix}1\\2\\2\end{pmatrix}, \begin{pmatrix}2\\5\\3\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}7\\7\\3\end{pmatrix}$

14. $W = Span\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}4\\4\\3\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}3\\5\\-5\end{pmatrix}$

15. $W = Span\left(\begin{pmatrix}1\\1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}2\\4\\5\\3\end{pmatrix}$

16. $W = Span\left(\begin{pmatrix}1\\1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\\1\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}1\\9\\3\\1\end{pmatrix}$

17. $W = Span\left(\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}1\\3\\5\\7\end{pmatrix}$

18. $W = Span\left(\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\\-1\\1\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}1\\3\\5\\7\end{pmatrix}$

19. $W = Span\left(\begin{pmatrix}3\\-1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\3\\-1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\3\\1\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}4\\2\\-6\\8\end{pmatrix}$

20. $W = Span\left(\begin{pmatrix}1\\2\\2\\0\end{pmatrix}, \begin{pmatrix}2\\-1\\0\\2\end{pmatrix}, \begin{pmatrix}2\\0\\-1\\-2\end{pmatrix}\right), \boldsymbol{u} = \begin{pmatrix}7\\7\\3\\1\end{pmatrix}$

In exercises 21 - 27 answer true or false and give an explanation.

21. If $W = Span(\boldsymbol{w}_1, \boldsymbol{w}_2)$ is a **subspace** of $\mathbb{R}^4$ and $\boldsymbol{w}_1 \cdot \boldsymbol{u} = \boldsymbol{w}_2 \cdot \boldsymbol{u} = 1$ then $\boldsymbol{u} \in W^{\perp}$.

22. If W is a **subspace** of $\mathbb{R}^5$ and $(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ is a basis of $W^{\perp}$ then $dim(W) = 3$.

23. If W is a **subspace** of $\mathbb{R}^n$ and $\boldsymbol{u} \in W$ then $Proj_W(\boldsymbol{u}) = \boldsymbol{u}$.

24. If W is a **subspace** of $\mathbb{R}^n$, $\boldsymbol{u} \in \mathbb{R}^n$ and $Proj_W(\boldsymbol{u}) = 0$ then $\boldsymbol{u} \in W$.

25. If W is a **subspace** of $\mathbb{R}^n$, $\boldsymbol{u} \in \mathbb{R}^n$ and $Proj_W(\boldsymbol{u}) = \boldsymbol{u}$ then $\boldsymbol{u} \in W$.

26. If W is a **subspace** of $\mathbb{R}^n$ and the vector $\boldsymbol{u}$ belongs to W and $W^{\perp}$ then $\boldsymbol{u} = \boldsymbol{0}_n$.

27. If $\boldsymbol{u} = \boldsymbol{w} + \boldsymbol{z}$ where $\boldsymbol{w}$ is in a **subspace** W of $\mathbb{R}^n$ then $\boldsymbol{w} = Proj_W(\boldsymbol{u})$.

Challenge Exercises (Problems)

1. Find a formula that gives the **distance** of the point (h, k) from the line with equation $ax + by = c$. (Hint: Treat the point (h, k) as the vector $\begin{pmatrix} h \\ k \end{pmatrix}$ and the line as the **affine subspace** $Span\left(\begin{pmatrix} -b \\ a \end{pmatrix}\right) + \boldsymbol{p}$ where $\boldsymbol{p} = \begin{pmatrix} \frac{c}{a} \\ 0 \end{pmatrix}$ if $a \neq 0$ and $\boldsymbol{p} = \begin{pmatrix} 0 \\ \frac{c}{b} \end{pmatrix}$ if $a = 0$).

2. Let a, b, c be non-zero real numbers. Find a formula that gives the **distance** of the point (h, k, l) in $\mathbb{R}^3$ from the plane with equation $ax + by + cz = d$. (Hint: Express the plane as an **affine subspace** $W + \boldsymbol{p}$ where W is an appropriate two dimensional **subspace** of $\mathbb{R}^3$).

3. Let W be a **subspace** of $\mathbb{R}^n$. Assume $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W, \boldsymbol{z}_1, \boldsymbol{z}_2 \in W^{\perp}$ and $\boldsymbol{w}_1 + \boldsymbol{z}_1 = \boldsymbol{w}_2 + \boldsymbol{z}_2$. Prove that $\boldsymbol{w}_1 = \boldsymbol{w}_2$ and $\boldsymbol{z}_1 = \boldsymbol{z}_2$.

4. Let W be a **subspace** of $\mathbb{R}^n$ and $\boldsymbol{u}$ an $n-$vector. Prove that $\boldsymbol{u} \in W^{\perp}$ if and only if $Proj_W(\boldsymbol{u}) = \boldsymbol{0}_n$.

5. Let E be a **symmetric** $n \times n$ matrix which satisfies $E^2 = E$. Set $F = I_n - E$.

a) Prove that F is **symmetric** and $F^2 = F$.

b) Prove that $EF = FE = \boldsymbol{0}_{n \times n}$.

c) Prove $col(E) \cap col(F) = \{\boldsymbol{0}_n\}$ and $col(E) + col(F) = \mathbb{R}^n$. Use this to conclude that $rank(E) + rank(F) = n$.

d) Prove that $col(F) = col(E)^{\perp}$.

e) Set $W = col(E)$. Prove that E is the **standard matrix** of $Proj_W$.

6. Let X and Y be **subspaces** of $\mathbb{R}^n$. Prove that

$$(Proj_Y \circ Proj_X)(\boldsymbol{u}) = Proj_Y(Proj_X(\boldsymbol{u})) = \mathbf{0}_n$$

for every vector $\boldsymbol{u}$ in $\mathbb{R}^n$ if and only if $X \perp Y$.

7. Let W be a **subspace** of $\mathbb{R}^n$ and $\boldsymbol{u}$ a vector in $\mathbb{R}^n$. Prove that $\parallel Proj_W(\boldsymbol{u}) \parallel \leq \parallel \boldsymbol{u} \parallel$ with equality if and only if $\boldsymbol{u} \in W$.

8. Let W be a **subspace** of $\mathbb{R}^n$ and $\boldsymbol{u}$ a vector in $\mathbb{R}^n$. Prove that $dist(\boldsymbol{u}, W) \leq \parallel \boldsymbol{u} \parallel$ with equality if and only if $\boldsymbol{u} \in W^\perp$.

Quiz Solutions

1. Set $\boldsymbol{v}_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$, $\boldsymbol{v}_2 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$ and $\boldsymbol{v}_3 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$ and $\mathcal{O} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$. Then $\mathcal{O}$ is the desired **orthonormal basis** of W. Not right, see **Method** (8.2.1).

2. The **orthogonal complement** to W is the **solution space** to the **homogeneous linear system** with **coefficient matrix** A^{Tr} where A is the matrix with the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ as its columns:

$$A^{Tr} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

The **reduced echelon form** of A^{Tr}, obtained using **Gaussian elimination**, is

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

The **null space** of R, hence A^{Tr}, is $Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}\right)$. Thus, $\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$ is a **unit vector** in $W^\perp$. Not right, see **Method** (2.6.1).

3. $\begin{pmatrix} 2 & 4 & 2 \\ 0 & 2 & -2 \\ 0 & 0 & 2 \end{pmatrix}$. Not right, see **Method** (5.4.3).

4. $Proj_{\boldsymbol{v}_1}(\boldsymbol{u}) = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix}, Proj_{\boldsymbol{v}_2}(\boldsymbol{u}) = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$ and $Proj_{\boldsymbol{v}_3}(\boldsymbol{u}) = \mathbf{0}_4$.

$\boldsymbol{u} - Proj_{\boldsymbol{v}_1}(\boldsymbol{u}) - Proj_{\boldsymbol{v}_2}(\boldsymbol{u}) - Proj_{\boldsymbol{v}_3}(\boldsymbol{u}) = \begin{pmatrix} \frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \end{pmatrix}$

Not right, see **Method** (2.6.3).

8.4. Diagonalization of Real Symmetric Matrices

In this section we consider a real $n \times n$ **symmetric matrix** A and demonstrate that A is **similar** to a **diagonal matrix** via an **orthogonal matrix** . That is, there is an **orthogonal matrix** Q such that $Q^{Tr}AQ$ is a **diagonal matrix**. Equivalently, there exists an **orthonormal basis** of $\mathbb{R}^n$ consisting of **eigenvectors** for A.

Am I Ready for This Material

Readiness Quiz

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

In this section it is essential to have a mastery of the following concepts:

symmetric matrix

diagonal matrix

similar matrices

eigenvector of a square matrix

eigenvalue of a square matrix

orthogonal or perpendicular vectors

orthogonal basis for a **subspace** of $\mathbb{R}^n$

orthonormal basis for a **subspace** of $\mathbb{R}^n$

orthogonal matrix

field of complex numbers

norm of a complex number

conjugate of a complex number

Quiz

1. Find an **orthonormal basis** for the **subspace** $W = Span\left(\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}\right)$ of $\mathbb{R}^3$.

2. Find an **orthonormal basis** for the **subspace** $w = Span\left(\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\right)$ of $\mathbb{R}^3$.

3. Show that the matrix $A = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$ is **diagonalizable** by finding a **basis** of $\mathbb{R}^2$ consisting of **eigenvectors** for A.

4. Show that the matrix $A = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$ is **diagonalizable** by finding a **basis** of $\mathbb{R}^3$ consisting of **eigenvectors** for A.

Quiz Solutions

New Concepts

In this section we introduce two new concepts:

orthogonally diagonalizable matrix

spectral decomposition of a **symmetric matrix**.

Theory (Why It Works)

In section (7.2) we obtained a criterion for a real matrix A to be **similar** to a real **diagonal matrix** D. Such matrices were called **diagonalizable** and an **invertible matrix** P such that $P^{-1}AP = D$ was called a **diagonalizing matrix**.

The main criterion for a real $n \times n$ matrix A to be **diagonalizable** is that there exist a **basis** of $\mathbb{R}^n$ consisting of **eigenvectors** for A. In turn, this requires that all of the roots of the **characteristic polynomial**, $\chi_A(\lambda)$, of A (which are the **eigenvalues** of A) be real and for each **eigenvalue** α the **geometric multiplicity** of α must be equal to the **algebraic multiplicity** of α.

Ordinarily, it is not possible to predict whether a real $n \times n$ matrix is **diagonalizable** without using the algorithm developed in section (7.2). However, as we shall see below there is a large class of real square matrices which are **diagonalizable**, namely the **symmetric matrices**, that is, those matrices A which are equal to their own **transpose**: $A^{Tr} = A$. We start with some examples.

Example 8.4.1. Let $A = \begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}$.

The **characteristic polynomial** of A is $\chi_A(\lambda) = det(A - \lambda I_2) =$

$$det \begin{pmatrix} 6-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix} =$$

$$(6-\lambda)(3-\lambda) - 4 = \lambda^2 - 9\lambda + 14 = (\lambda-2)(\lambda-7) \tag{8.41}$$

Since the roots of $\chi_A(\lambda)$ are real and simple (the **algebraic multiplicity** of each is one) we can immediately conclude that the matrix A is **diagonalizable**. We find a **diagonalizing matrix** by finding an **eigenvector** for each of the **eigenvalues**.

Finding an eigenvector for the eigenvalue 2:

$A-2I_2 = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{pmatrix}$. The **null space** of this matrix is **spanned** by $\begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix}$. We multiply by two in order to "clear fractions" and so use the vector $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

Finding an eigenvector for the eigenvalue 7:

$A - 7I_2 = \begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix}$. This matrix has **reduced echelon form** $\begin{pmatrix} 1 & -2 \\ 0 & 0 \end{pmatrix}$. The **null space** of this matrix is **spanned** by $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Set $P = \begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix}$. Then $P^{-1}AP = \begin{pmatrix} 2 & 0 \\ 0 & 7 \end{pmatrix}$. Notice that the two columns of P are **orthogonal**:

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

We shall see that this is to be expected and always occurs for **eigenvectors** with distinct **eigenvalues** of a **symmetric matrix**. Since these vectors are **orthogonal**, we can **normalize** them to obtain an **orthonormal basis** for $\mathbb{R}^2$ consisting of **eigenvectors** for A and a **orthogonal matrix** Q which is a **diagonalizing matrix for** A,

$$Q = \begin{pmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}.$$

Example 8.4.2. Let $A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 2 & 2 & -1 \end{pmatrix}$. The **characteristic polynomial** of A is

$$\chi_A(\lambda) = det(A - \lambda I_3) = det \begin{pmatrix} 1-\lambda & 0 & 2 \\ 0 & 1-\lambda & 2 \\ 2 & 2 & -1-\lambda \end{pmatrix} =$$

$$(1-\lambda)(1-\lambda)(-1-\lambda) - 4(1-\lambda) - 4(1-\lambda) =$$

$$(\lambda - 1)(9 - \lambda^2) = -(\lambda-1)(\lambda-3)(\lambda+3) \tag{8.42}$$

So, all the roots of $\chi_A(\lambda)$ are real and simple and therefore A is **diagonalizable**. We find a **basis** of **eigenvectors** using **Method** (7.1.2):

Finding an eigenvector for the eigenvalue 1:

$A - I_3 = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 2 \\ 2 & 2 & -2 \end{pmatrix}$. The **reduced echelon form** of $A - I_3$ is $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. The **null space** of this matrix is **spanned** by $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$

Finding an eigenvector for the eigenvalue 3:

$A - 3I_3 = \begin{pmatrix} -2 & 0 & 2 \\ 0 & -2 & 2 \\ 2 & 2 & -4 \end{pmatrix}$. The **reduced echelon form** of $A - 3I_3$ is $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$.

The **null space** of $A - 3I_3$ is **spanned** by $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Finding an eigenvector for the eigenvalue -3:

$A + 3I_3 = \begin{pmatrix} 4 & 0 & 2 \\ 0 & 4 & 2 \\ 2 & 2 & 2 \end{pmatrix}$. The **reduced echelon form** of $A + 3I_3$ is $\begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}$. The **null space** of this matrix is **spanned** by $\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$.

The sequence of vectors $\mathcal{B} = \left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right)$ is a **basis** of **eigenvectors** for A. The matrix P whose columns are these vectors is a **diagonalizing matrix** for A. Notice that this is an **orthogonal basis** for $\mathbb{R}^3$ since the vectors are mutually **orthogonal**:

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

By **normalizing** the vectors of the **basis** $\mathcal{B}$ we get an **orthonormal basis**, $\mathcal{O}$, consisting of **eigenvectors** for A:

$$\left(\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}\right)$$

The **orthogonal matrix** $Q = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{pmatrix}$ is a **diagonalizing matrix**.

The phenomenon observed in **Example** (8.4.1) and **Example** (8.4.2), where **eigenvectors** corresponding to distinct **eigenvalues** are **orthogonal** is not a coincidence; rather it is a consequence of the fact that the matrix A is **symmetric**. We will next prove a theorem which implies this but has greater generality and will later be useful in demonstrating that all the **eigenvalues** of a real **symmetric matrix** are all real.

Theorem 8.4.1. *Let A be a real $n \times n$ **symmetric matrix** and let $\alpha \neq \beta$ be complex **eigenvalues** of A with respective complex **eigenvectors** $\boldsymbol{z}$ and $\boldsymbol{w}$. Then $\boldsymbol{z} \cdot \boldsymbol{w} = 0$.*

Proof. Here $\boldsymbol{z} \cdot \boldsymbol{w}$ has the usual meaning: If $\boldsymbol{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$ and $\boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ then $\boldsymbol{z} \cdot \boldsymbol{w} = z_1 w_1 + z_2 w_2 + \cdots + z_n w_n$. Consider the **dot product** of $\boldsymbol{z}$ with $A\boldsymbol{w}$. Since $A\boldsymbol{w} = \beta\boldsymbol{w}$ this is just $\beta(\boldsymbol{z} \cdot \boldsymbol{w})$. On the other hand, $\boldsymbol{z} \cdot (A\boldsymbol{w}) = \boldsymbol{z}^{Tr}(A\boldsymbol{w})$. Since $\boldsymbol{z}^{Tr}(A\boldsymbol{w})$ is a scalar it is equal to its transpose:

$$\boldsymbol{z}^{Tr}(A\boldsymbol{w}) = [\boldsymbol{z}^{Tr}(A\boldsymbol{w})]^{Tr} = (A\boldsymbol{w})^{Tr}(\boldsymbol{z}^{Tr})^{Tr} = (\boldsymbol{w}^{Tr}A^{Tr})\boldsymbol{z}.$$

By the **associativity of matrix multiplication** $(\boldsymbol{w}^{Tr}A^{Tr})\boldsymbol{z} = \boldsymbol{w}^{Tr}(A^{Tr}\boldsymbol{z})$. Since A is **symmetric**, $A^{Tr} = A$. Therefore $\boldsymbol{w}^{Tr}(A^{Tr}\boldsymbol{z}) = \boldsymbol{w}^{Tr}(A\boldsymbol{z})$. Since $\boldsymbol{z}$ is an **eigenvector** with **eigenvalue** α we have $\boldsymbol{w}^{Tr}(A\boldsymbol{z}) = \boldsymbol{w}^{Tr}(\alpha\boldsymbol{z}) = \alpha(\boldsymbol{w} \cdot \boldsymbol{z})$. Thus,

$$\beta(\boldsymbol{z} \cdot \boldsymbol{w}) = \alpha(\boldsymbol{w} \cdot \boldsymbol{z}) = \alpha(\boldsymbol{z} \cdot \boldsymbol{w})$$

It then follows that $(\beta - \alpha)(\boldsymbol{z} \cdot \boldsymbol{w}) = 0$. Since $\alpha \neq \beta, \beta - \alpha \neq 0$ and it must be the case that $\boldsymbol{z} \cdot \boldsymbol{w} = 0$ as claimed. $\square$

Theorem (8.4.1) says nothing about **orthogonality** when an **eigenvalue** has **geometric multiplicity** greater than one, which can occur as the next example illustrates.

Example 8.4.3. Let $A = \begin{pmatrix} 0 & -4 & -2 \\ -4 & 0 & 2 \\ -2 & 2 & -3 \end{pmatrix}$. This matrix is **symmetric**. We find the **characteristic polynomial** and the **eigenvalues** .

$$\chi_A(\lambda) = det(A - \lambda I_3) = det \begin{pmatrix} -\lambda & -4 & -2 \\ -4 & -\lambda & 2 \\ -2 & 2 & -3-\lambda \end{pmatrix} =$$

$$(-\lambda)(-\lambda)(-3-\lambda)+16+16-(-\lambda)(-2)(2)-(-\lambda)(-2)(-2)-(-3-\lambda)(-4)(-4) =$$

$$-\lambda^3 - 3\lambda^2 + 24\lambda + 80 = -(\lambda+4)^2(\lambda-5)$$

So $\chi_A(\lambda)$ has -4 as a double root and 5 as a simple root. We find a **basis** for each of the corresponding **eigenspaces**.

Finding an eigenvector for the eigenvalue 5:

$A - 5I_3 = \begin{pmatrix} -5 & -4 & -2 \\ -4 & -5 & 2 \\ -2 & 2 & -8 \end{pmatrix}$. The **reduced echelon form** of $A - 5I_3$ is $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$.

The **null space** of this matrix is **spanned** by $\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$

Finding a basis for the eigenspace for the eigenvalue -4:

$A + 4I_3 = \begin{pmatrix} 4 & -4 & -2 \\ -4 & 4 & 2 \\ -2 & 2 & 1 \end{pmatrix}$. The matrix $A + 4I_3$ has **reduced echelon form** $\begin{pmatrix} 1 & -1 & -\frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Notice that $A+4I_3$ has **rank** one and **nullity** two and therefore A is **diagonalizable**. We find a **basis** for $null(A + 4I_3) = E_{-4}$. After adjoining this basis to $\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$ we obtain a **basis** for $\mathbb{R}^3$ consisting of **eigenvectors** for A.

The **null space** of $A + 4I_3$ is **spanned** by $\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \right)$. Notice that each of these vectors is **orthogonal** to the **eigenvector** $\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$ for the **eigenvalue** 5 as predicted by **Theorem** (8.4.1).

The vectors $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ are not **orthogonal**. However, we can apply the Gram-Schmidt process, **Method** (8.2.1), to obtain an **orthogonal basis** for $null(A + 4I_3)$. Specifically, we replace $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ with its **projection orthogonal** to $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. The vector we obtain is $\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 2 \end{pmatrix}$. We can clear fractions by multiplying by two to get the vector $\begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$. We therefore have the following **orthogonal basis** consisting of **eigenvectors** for A:

$$\mathcal{B} = \left(\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix} \right)$$

After **normalizing** the vectors of $\mathcal{B}$ we get the **orthonormal basis**

$$\mathcal{O} = \left(\begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{pmatrix} \right).$$

The matrix $P = \begin{pmatrix} -\frac{2}{3} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \\ \frac{1}{3} & 0 & \frac{1}{\sqrt{18}} \end{pmatrix}$ is an **orthogonal matrix** and a **diagonalizing matrix** for A.

Definition 8.10. A real $n \times n$ matrix is ***orthogonally diagonalizable*** if there is an $n \times n$ **orthogonal matrix** Q such that $Q^{-1}AQ$ is a **diagonal matrix**.

An immediate criterion for a square matrix to be **orthogonally diagonalizable** is given by the following theorem:

Theorem 8.4.2. *An $n \times n$ matrix A is* **orthogonally diagonalizable** *if and only if there exists an* **orthonormal basis** *for $\mathbb{R}^n$ consisting of* **eigenvectors** *for A.*

It is our ultimate goal in this section to show that the **orthogonally diagonalizable** matrices are the **symmetric matrices**, in particular, that symmetric matrices are diagonalizable. Towards that end we first show that all the **eigenvalues** of a real **symmetric matrix** are real.

Theorem 8.4.3. *Let A be a real $n \times n$* **symmetric matrix**. *Then every* **eigenvalue** *of A is real.*

Proof. Let α be complex **eigenvalue** of A. We need to show that α is real, or equivalently, that $\overline{\alpha} = \alpha$. Suppose to the contrary that $\alpha \neq \overline{\alpha}$.

Let $\boldsymbol{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$ be a complex **eigenvector** of A with **eigenvalue** α. Since A is a real matrix it follows from **Theorem** (7.3.3) that $\overline{\boldsymbol{z}} = \begin{pmatrix} \overline{z_1} \\ \overline{z_2} \\ \vdots \\ \overline{z_n} \end{pmatrix}$ is an **eigenvector** of A with **eigenvalue** $\overline{\alpha}$. Since A is **symmetric**, we can invoke **Theorem** (8.4.1). Since $\alpha \neq \overline{\alpha}$ we can conclude that $\boldsymbol{z} \cdot \overline{\boldsymbol{z}} = 0$. However,

$$\boldsymbol{z} \cdot \overline{\boldsymbol{z}} = z_1\overline{z_1} + z_2\overline{z_2} + \cdots + z_n\overline{z_n} = \parallel z_1 \parallel^2 + \parallel z_2 \parallel^2 + \cdots + \parallel z_n \parallel^2 .$$

Then $z_1 = z_2 = \cdots = z_n = 0$ and $\boldsymbol{z} = \mathbf{0}_n$ contradicting the assumption that $\boldsymbol{z}$ is an **eigenvector**. Thus, $\alpha = \overline{\alpha}$ and α is real. □

We can now prove our main theorem, that a matrix A is **orthogonally diagonalizable** if and only if A is **symmetric**. This is known as the **Real Spectral Theorem**.

Theorem 8.4.4. *An $n \times n$ real matrix A is* **orthogonally diagonalizable** *if and only if A is* **symmetric**.

Proof. One direction is easy: If a matrix A is **orthogonally diagonalizable** then A is **symmetric**. This depends on the fact that for an **orthogonal matrix** $Q, Q^{-1} = Q^{Tr}$.

Suppose that A is **orthogonally diagonalizable** and that Q is an **orthogonal matrix** such that $Q^{-1}AQ = D$. Then $A = QDQ^{-1}$. As noted, $Q^{-1} = Q^{Tr}$ and therefore $A = QDQ^{Tr}$. Then $A^{Tr} = (QDQ^{Tr})^{Tr} = (Q^{Tr})^{Tr}D^{Tr}Q^{Tr}$. For all matrices $X, (X^{Tr})^{Tr} = X$, in particular, $(Q^{Tr})^{Tr} = Q$. Since D is a **diagonal matrix**, D is **symmetric**. Consequenlty, $A^{Tr} = QDQ^{Tr} = A$ and so A is **symmetric**.

The converse, that a **symmetric matrix** is **orthogonally diagonalizable**, is significantly more difficult. The proof is by **induction** on n. We begin with $n = 2$.

Let A be a 2×2 **symmetric matrix**. We know from **Theorem** (8.4.3) that the **eigenvalues** of A are real. Let α_1, α_2 be the **eigenvalues** of A. Suppose first that $\alpha_1 \neq \alpha_2$ and let $\boldsymbol{x}_1$ be an **eigenvector** for A with **eigenvalue** α_1 and $\boldsymbol{x}_2$ an **eigenvector** for A with **eigenvalue** α_2. By **Theorem** (8.4.1), $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are **orthogonal**. Replacing $\boldsymbol{x}_i$, if necessary, with $\boldsymbol{x}_i' = \frac{1}{\|\boldsymbol{x}_i\|}\boldsymbol{x}_i$, we obtain an **orthonormal basis** for $\mathbb{R}^2$ consisting of **eigenvectors** for A.

Assume that $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ has a unique **eigenvalue** α with **algebraic multiplicity** two. The **characteristic polynomial** of A is $\lambda^2 - (a+c)\lambda + (ac - b^2)$. In order for this polynomial to have two identical roots its **discriminant** must equal zero. The discriminant of a quadratic $Ax^2 + Bx + C$ is $\Delta = B^2 - 4AC$. Therefore, we must have $[-(a+c)]^2 - 4(ac - b^2) = 0$. However,

$$[-(a+c)]^2 - 4(ac-b^2) = a^2 + 2ac + c^2 - 4ac + 4b^2 =$$

$$a^2 - 2ac + c^2 + 4b^2 = (a-c)^2 + (2b)^2 = 0$$

For the sum of two (real) squares to be zero, each number must be zero. So, $b = 0$ and $a = c$ and therefore the matrix $A = aI_2 = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ and the **standard basis** for $\mathbb{R}^2$ is an **orthonormal basis** consisting of **eigenvectors** for A.

We now have to do the inductive case. We therefore assume that if B is a real **symmetric** $(n-1)\times(n-1)$ matrix then there exists an $(n-1)$ **orthogonal matrix** P such that $P^{Tr}AP$ is a **diagonal matrix**. We must show that the result can be extended to real $n \times n$ **symmetric matrices**.

Let A be a real $n\times n$ **symmetric matrix**. By **Theorem** (8.4.3) the **eigenvalues** of A are real. Let α be an **eigenvalue** of A and $\boldsymbol{x}$ a **eigenvector** with eigenvalue α. If necessary, we can **normalize** $\boldsymbol{x}$ and assume that $\boldsymbol{x}$ has **norm** one. Set $\boldsymbol{x}_1 = \boldsymbol{x}$ and extend $\boldsymbol{x}_1$ to an **orthonormal basis** $\mathcal{O} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$. That this can be done is a consequence of the fact that every **linearly independen**t sequence can be extended to a **basis** for $\mathbb{R}^n$ and the Gram-Schmidt process, **Method** (8.2.1).

Let $Q_1 = (\boldsymbol{x}_1\ \boldsymbol{x}_2 \ldots\ \boldsymbol{x}_n)$, an **orthogonal matrix**. Consider the matrix $AQ_1 = (A\boldsymbol{x}_1\ A\boldsymbol{x}_2 \ldots\ A\boldsymbol{x}_n)$. Since $\boldsymbol{x}_1$ is an **eigenvector** of A with **eigenvalue** α we have

$$AQ_1 = (\alpha\boldsymbol{x}_1\ A\boldsymbol{x}_2 \ldots\ A\boldsymbol{x}_n)$$

Then

$$Q_1^{Tr}AQ_1 = Q_1^{Tr}(\alpha\boldsymbol{x}_1\ A\boldsymbol{x}_2 \ldots\ A\boldsymbol{x}_n) = (Q_1^{Tr}(\alpha\boldsymbol{x}_1)\ Q_1^{Tr}(A\boldsymbol{x}_2) \ldots\ Q_1^{Tr}(A\boldsymbol{x}_n)) =$$

$$(\alpha(Q_1^{Tr}\boldsymbol{x}_1)\ Q_1^{Tr}(A\boldsymbol{x}_2) \ldots\ Q_1^{Tr}(A\boldsymbol{x}_n))$$

Since $Q_1^{Tr} = Q_1^{-1}, Q_1^{Tr}Q_1 = I_n$. On the other hand

$$Q_1^{Tr}Q_1 = Q_1^{Tr}(\boldsymbol{x}_1\ \boldsymbol{x}_2 \ldots\ \boldsymbol{x}_n) = (Q_1^{Tr}\boldsymbol{x}_1\ Q_1^{Tr}\boldsymbol{x}_2 \ldots\ Q_1^{Tr}\boldsymbol{x}_n).$$

The point of this is that $Q_1^{Tr}\boldsymbol{x}_1 = \boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

Therefore the first column of $Q_1^{Tr}AQ_1$ is $\begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

Next notice that $Q_1^{Tr}AQ_1$ is **symmetric** since A is symmetric:

$$[Q_1^{Tr}AQ_1]^{Tr} = Q_1^{Tr}A^{Tr}(Q_1^{Tr})^{Tr} = Q_1^{Tr}AQ_1$$

Therefore the first row of $Q_1^{Tr}AQ_1$ must be $(\alpha\ 0 \ldots\ 0)$ and hence the matrix $Q_1^{Tr}AQ_1$ has the form $\begin{pmatrix} \alpha & \mathbf{0}_{n-1}^{Tr} \\ \mathbf{0}_n & B \end{pmatrix}$ where B is a real $(n-1) \times (n-1)$ **symmetric matrix**.

By the inductive hypothesis there is an $n-1 \times n-1$ **orthogonal matrix** P such that $P^{Tr}BP = diag(\alpha_2, \alpha_3, \ldots, \alpha_n)$, a diagonal matrix.

Set $Q_2 = \begin{pmatrix} 1 & \mathbf{0}_{n-1}^{Tr} \\ \mathbf{0}_{n-1} & P \end{pmatrix}$. Then Q_2 is an $n \times n$ **orthogonal matrix**. Now let $Q = Q_1Q_2$. Since Q is a product of **orthogonal matrices**, Q is an orthogonal matrix. We claim that $Q^{Tr}AQ$ is a diagonal matrix.

$$Q^{Tr}AQ = (Q_1Q_2)^{Tr}A(Q_1Q_2) =$$

$$Q_2^{Tr}(Q_1A^{Tr}Q_1)Q_2 = Q_2^{Tr}\begin{pmatrix} \alpha & \mathbf{0}_{n-1}^{Tr} \\ \mathbf{0}_{n-1} & B \end{pmatrix} Q_2 =$$

$$\begin{pmatrix} 1 & \mathbf{0}_{n-1}^{Tr} \\ \mathbf{0}_{n-1} & P^{Tr} \end{pmatrix} \begin{pmatrix} \alpha & \mathbf{0}_{n-1}^{Tr} \\ \mathbf{0}_{n-1} & B \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}_{n-1}^{Tr} \\ \mathbf{0}_{n-1} & P \end{pmatrix} =$$

$$\begin{pmatrix} \alpha & \mathbf{0}_{n-1}^{Tr} \\ \mathbf{0}_{n-1} & P^{Tr}BP \end{pmatrix} = \begin{pmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{pmatrix}$$

This completes the theorem □

Let A be a real $n \times n$ **symmetric matrix** and Q an **orthogonal matrix** such that $Q^{Tr}AQ = D = diag(\alpha_1, \alpha_2, \dots, \alpha_n) = \begin{pmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_n \end{pmatrix}$. Then $A = QDQ^{Tr}$. Let the columns of Q be $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$. Then

$$A = (\boldsymbol{x}_1\ \boldsymbol{x}_2\ \dots\ \boldsymbol{x}_n) \begin{pmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_n \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1^{Tr} \\ \boldsymbol{x}_2^{Tr} \\ \vdots \\ \boldsymbol{x}_n^{Tr} \end{pmatrix}$$

$(\boldsymbol{x}_1\ \boldsymbol{x}_2\ \dots\ \boldsymbol{x}_n) \begin{pmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_n \end{pmatrix} = (\alpha_1\boldsymbol{x}_1\ \alpha_2\boldsymbol{x}_2\ \dots\ \alpha_n\boldsymbol{x}_n)$ and consequently

$$A = (\alpha_1\boldsymbol{x}_1\ \alpha_2\boldsymbol{x}_2\ \dots\ \alpha_n\boldsymbol{x}_n) \begin{pmatrix} \boldsymbol{x}_1^{Tr} \\ \boldsymbol{x}_2^{Tr} \\ \vdots \\ \boldsymbol{x}_n^{Tr} \end{pmatrix} = \alpha_1\boldsymbol{x}_1\boldsymbol{x}_1^{Tr} + \alpha_2\boldsymbol{x}_2\boldsymbol{x}_2^{Tr} + \dots + \alpha_n\boldsymbol{x}_n\boldsymbol{x}_n^{Tr}.$$

Each of the matrices $\boldsymbol{x}_i\boldsymbol{x}_i^{Tr}$ is a **rank** one matrix and **symmetric**. The scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ are the **eigenvalues** of the matrix A and the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ are the corresponding **eigenvectors** of A. This decomposition has a special name.

Definition 8.11. An expression of the form $A = \alpha_1 \boldsymbol{x}_1 \boldsymbol{x}_1^{Tr} + \alpha_2 \boldsymbol{x}_2 \boldsymbol{x}_2^{Tr} + \cdots + \alpha_n \boldsymbol{x}_n \boldsymbol{x}_n^{Tr}$ for a symmetric matrix A where $(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$ is an orthonormal basis of $\mathbb{R}^n$ is called a ***spectral decomposition of*** A.

Example 8.4.4. Find a spectral decomposition for the matrix A of Example (8.4.2)

We determined that the eigenvalues of A are 1, 3 and -3 with corresponding eigenvectors $\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$ and $\begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$.

This leads to the following spectral decomposition:

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

The spectral decomposition can be used to construct a symmetric matrix given its eigenvalues and corresponding eigenvectors (provided the latter is an orthonormal basis for $\mathbb{R}^n$).

Example 8.4.5. Find a 3×3 symmetric matrix which has eigenvalues 3, -6, 4 with corresponding eigenvectors $\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

These vectors are mutually orthogonal. We normalize to obtain an orthonormal basis:

$$\boldsymbol{x}_1 = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} \frac{4}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

The required matrix is

$$A = 3\boldsymbol{x}_1 \boldsymbol{x}_1^{Tr} - 6\boldsymbol{x}_2 \boldsymbol{x}_2^{Tr} + 4\boldsymbol{x}_3 \boldsymbol{x}_3^{Tr} =$$

$$3 \begin{pmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{pmatrix} - 6 \begin{pmatrix} \frac{16}{18} & -\frac{4}{18} & -\frac{4}{18} \\ \frac{-4}{18} & \frac{1}{18} & \frac{1}{18} \\ \frac{-4}{18} & \frac{1}{18} & \frac{1}{18} \end{pmatrix} + 4 \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} =$$

$$\begin{pmatrix} -5 & 2 & 2 \\ 2 & 3 & -1 \\ 2 & -1 & 3 \end{pmatrix}$$

What You Can Now Do

1. Given an $n \times n$ **symmetric matrix** A find an **orthonormal basis** for $\mathbb{R}^n$ consisting of **eigenvectors** for A.

2. Given an $n \times n$ **symmetric matrix** A find an **orthogonal matrix** Q and **diagonal matrix** D such that $Q^{Tr}AQ = D$.

3. Given an $n \times n$ **symmetric matrix** A find a **spectral decomposition** for A.

4. Given an **orthogonal basis** $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ for $\mathbb{R}^n$ and scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ find a **symmetric matrix** A such that $\boldsymbol{w}_i$ is an **eigenvector** of A with **eigenvalue** α_i for $1 \leq i \leq n$.

Method (How To Do It)

Method 8.4.1. Given an $n \times n$ **symmetric matrix** A find an **orthonormal basis** for $\mathbb{R}^n$ consisting of **eigenvectors** for A.

Use **Method** (7.1.1) to find the **characteristic polynomial** and determine the **eigenvalues** of A. For each **eigenvalue** α of A use **Method** (7.1.3) to find a **basis** $\mathcal{B}_\alpha$ for the **eigenspace** E_α. If all the **eigenvalues** are simple (have **algebraic multiplicity** one) then the sequence of **eigenvectors** obtained will be an **orthogonal basis** for $\mathbb{R}^n$. If necessary, **normalize** the vectors of the sequence to obtain an **orthonormal basis**.

If some α has **algebraic multiplicity** (which is equal to its **geometric multiplicity**) greater than one apply the Gram-Schmidt process, **Method** (8.2.1), to the basis $\mathcal{B}_\alpha$ to obtain an **orthonormal basis** $\mathcal{O}_\alpha$ for E_α. The union of all the $\mathcal{O}_\alpha$ will be an **orthogonal basis** of $\mathbb{R}^n$ consisting of **eigenvectors** of the matrix A. **Normalizing**, if necessary, any **eigenvectors** for which occur for **eigenvalues** with algebraic multiplicity 1 yields a **orthonormal basis**.

Example 8.4.6. Find an **orthonormal basis** for $\mathbb{R}^2$ consisting of **eigenvectors** for the **symmetric matrix** $A = \begin{pmatrix} -1 & 2 \\ 2 & 2 \end{pmatrix}$.

The **characteristic polynomial** of A is $\chi_A(\lambda) =$

$$(-1-\lambda)(2-\lambda) - 4 = \lambda^2 - \lambda - 6 = (\lambda+2)(\lambda-3)$$

The **eigenvalues** are -2 and 3. The vector $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ is an **eigenvector** for -2 and $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is an **eigenvector** with **eigenvalue** 3. These are **orthogonal**. After **normalizing** we obtain the **orthonormal basis**: $\left(\begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \right)$.

Example 8.4.7. Find an **orthonormal basis** for $\mathbb{R}^3$ consisting of **eigenvectors** for the **symmetric matrix** $A = \begin{pmatrix} 2 & -2 & -3 \\ -2 & 1 & -2 \\ -3 & -2 & 2 \end{pmatrix}$.

The **characteristic polynomial** of A is $\chi_A(\lambda) =$

$$(2-\lambda)(1-\lambda)(2-\lambda) - 12 - 12 - 4(2-\lambda) - 9(1-\lambda) - 4(2-\lambda) =$$

$$-(\lambda^3 - 5\lambda^2 - 9\lambda + 45) = -(\lambda+3)(\lambda-3)(\lambda-5).$$

The **eigenvalues** are -3, 3 and 5. The **eigenspaces** are:

$E_{-3} = Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$, $E_3 = Span\left(\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right)$, and $E_5 = Span\left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right)$.

These **eigenvectors** are mutually **orthogonal**. After **normalizing** we obtain the **orthonormal basis**:

$$\left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right)$$

Example 8.4.8. Find an **orthonormal basis** for $\mathbb{R}^3$ consisting of **eigenvectors** for the **symmetric matrix** $A = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}$.

The **characteristic polynomial** of A is $\chi_A(\lambda) =$

$$(3-\lambda)^3 + 1 + 1 - (3-\lambda) - (3-\lambda) - (3-\lambda) =$$

$$-(\lambda^3 - 9\lambda^2 + 24\lambda - 20) = -(\lambda-2)^2(\lambda-5).$$

The **eigenvalues** are 2 with multiplicity two and 5 with multiplicity one. By **Method** (7.1.3) we obtain a **basis** for each of the **eigenspaces** . These are:

$$\mathcal{B}_5 = \left(\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}\right), \mathcal{B}_2 = \left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\right).$$

Note that the **basis** vectors $\boldsymbol{x} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ for E_2 are not **orthogonal**. We therefore replace $\boldsymbol{y}$ with $Proj_{\boldsymbol{x}^\perp}(\boldsymbol{y}) = \boldsymbol{y} - \frac{\boldsymbol{y}\cdot\boldsymbol{x}}{\boldsymbol{x}\cdot\boldsymbol{x}}\boldsymbol{x} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{-1}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}$.

Now the sequence $\left(\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix}\right)$ is an **orthogonal basis** of $\mathbb{R}^3$ consisting of **eigenvectors** for A. After **normalizing** we get the following **orthonormal basis** of $\mathbb{R}^3$ consisting of **eigenvectors** for A :

$$\begin{pmatrix} -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$$

Method 8.4.2. Given an $n \times n$ **symmetric matrix** A find an **orthogonal matrix** Q and **diagonal matrix** D such that $Q^{Tr}AQ = D$.

Use **Method** (8.4.1) to obtain an **orthonormal basis** $\mathcal{O} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ consisting of **eigenvectors** for A. Set $Q = (\boldsymbol{w}_1\ \boldsymbol{w}_2 \dots\ \boldsymbol{w}_n)$. Then

$$Q^{Tr}AQ = \begin{pmatrix} \alpha_1 & 0 & 0 & \dots & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_n \end{pmatrix}$$

where α_i is the **eigenvalue** for the **eigenvector** $\boldsymbol{w}_i$ ($A\boldsymbol{w}_i = \alpha_i \boldsymbol{w}_i$).

Example 8.4.9. Let $A = \begin{pmatrix} 5 & -3 \\ -3 & -3 \end{pmatrix}$. Find an **orthogonal matrix** Q such that $Q^{Tr}AQ$ is a **diagonal matrix**.

The **characteristic polynomial** of A is $\chi_A(\lambda) =$

$$\lambda^2 - 2\lambda - 24 = (\lambda + 4)(\lambda - 6)$$

The **eigenvalues** are -4 and 6. The **eigenspaces** are $E_{-4} = Span\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}\right)$ and $E_6 = Span\left(\begin{pmatrix} -3 \\ 1 \end{pmatrix}\right)$.

The sequence $\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \end{pmatrix}\right)$ is an **orthogonal basis** consisting of **eigenvectors** for A. We **normalize** to obtain the **orthonormal basis** $\left(\begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{pmatrix}, \begin{pmatrix} -\frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{pmatrix}\right)$. We make these the columns of a matrix Q :

$$Q = \begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix}$$

Then $Q^{Tr}AQ = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} 5 & -3 \\ -3 & -3 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix} = \begin{pmatrix} -4 & 0 \\ 0 & 6 \end{pmatrix}$.

Example 8.4.10. Let $A = \begin{pmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{pmatrix}$. Find an **orthogonal matrix** Q such that $Q^{Tr}AQ$ is a **diagonal matrix**.

The **characteristic polynomial** of A is $\chi_A(\lambda) =$

$$(1-\lambda)^3 - 8 - 8 - 4(1-\lambda) - 4(1-\lambda) - 4(1-\lambda) =$$

$$-\lambda^3 + 3\lambda^2 + 9\lambda - 27 = -(\lambda+3)(\lambda-3)^2.$$

The **eigenvalues** of A are -3 with multiplicity one and 3 with multiplicity two. Using **Method** (7.1.3) we obtain bases for the **eigenspaces**

$$\mathcal{B}_{-3} = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right), \mathcal{B}_3 = \left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right).$$

The **basis** vectors $\boldsymbol{x} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ for E_3 are not **orthogonal** so we need to use **Method** (8.2.1), to obtain an **orthogonal basis** for E_3. We do this by replacing $\boldsymbol{y}$ by $Proj_{\boldsymbol{x}^\perp}(\boldsymbol{y})$.

$$Proj_{\boldsymbol{x}^\perp}(\boldsymbol{y}) = \boldsymbol{y} - \frac{\boldsymbol{y}\cdot\boldsymbol{x}}{\boldsymbol{x}\cdot\boldsymbol{x}}\boldsymbol{x} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}.$$

Now the sequence $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}\right)$ is an **orthogonal basis** for $\mathbb{R}^3$ consisting of **eigenvectors** for A. We **normalize** the vectors of the sequence and, in this way, obtain the **orthonormal basis** $\left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}\right)$. Making these the columns of a matrix we get the **orthogonal matrix** Q where

$$Q = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$$

Then $Q^{Tr}AQ =$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} -3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Method 8.4.3. Given an $n \times n$ **symmetric matrix** A find a **spectral decomposition** for A.

Use **Method** (8.4.1) to obtain an **orthonormal basis** $(\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ consisting of **eigenvectors** for A. Assume that $A\boldsymbol{w}_i = \alpha_i \boldsymbol{w}_i$. Then

$$A = \alpha_1 \boldsymbol{w}_1 \boldsymbol{w}_1^{Tr} + \alpha_2 \boldsymbol{w}_2 \boldsymbol{w}_2^{Tr} + \dots + \alpha_n \boldsymbol{w}_n \boldsymbol{w}_n^{Tr}$$

is a **spectral decompositon** for A.

Example 8.4.11. Let $A = \begin{pmatrix} -1 & 2 \\ 2 & 2 \end{pmatrix}$ be the matrix of **Example** (8.4.6).

In **Example** (8.4.6) we found the **eigenvalues** of A are -2 and 3. We obtained the **orthonormal basis** $\left(\begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \right)$. The **spectral decomposition** we obtain is

$$A = (-2) \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}^{Tr} + 3 \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}^{Tr} =$$

$$(-2) \begin{pmatrix} \frac{4}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{1}{5} \end{pmatrix} + 3 \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{4}{5} \end{pmatrix}$$

Example 8.4.12. Let $A = \begin{pmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{pmatrix}$ the matrix of **Example** (8.4.10). There we determined that the **eigenvalues** of A are -3 with multiplicity one and 3 with multiplicity two. We also obtained the following **orthonormal basis** of $\mathbb{R}^3$ consisting of **eigenvectors** for A (with **eigenvalues** -3, 3, and 3, respectively):

$$\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}.$$

This gives the **spectral decomposition**

$$A = (-3)\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}^{Tr} + 3\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}^{Tr} + 3\begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}\begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}^{Tr} =$$

$$(-3)\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} + 3\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} + 3\begin{pmatrix} \frac{1}{6} & \frac{1}{6} & -\frac{2}{6} \\ \frac{1}{6} & \frac{1}{6} & -\frac{2}{6} \\ -\frac{2}{6} & -\frac{2}{6} & \frac{4}{6} \end{pmatrix}.$$

Since the **eigenvector** 3 has multiplicity two an **orthonormal basis** for E_3 is not unique (up to signs). In fact, $E_3 = Span\left(\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}\right)$.

This gives a different **orthonormal basis** of $\mathbb{R}^3$ consisting of **eigenvectors** for A :

$$\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}$$

From this **basis** we obtain the **spectral decomposition**:

$$A = (-3)\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}^{Tr} + 3\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}^{Tr} + 3\begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}\begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}^{Tr} =$$

$$(-3)\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} + 3\begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} + 3\begin{pmatrix} \frac{1}{6} & -\frac{2}{6} & \frac{1}{6} \\ -\frac{2}{6} & \frac{4}{6} & -\frac{2}{6} \\ \frac{1}{6} & -\frac{2}{6} & \frac{1}{6} \end{pmatrix}$$

These are different but note that

$$3\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} + 3\begin{pmatrix} \frac{1}{6} & \frac{1}{6} & -\frac{2}{6} \\ \frac{1}{6} & \frac{1}{6} & -\frac{2}{6} \\ -\frac{2}{6} & -\frac{2}{6} & \frac{4}{6} \end{pmatrix} = 3\begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} + 3\begin{pmatrix} \frac{1}{6} & -\frac{2}{6} & \frac{1}{6} \\ -\frac{2}{6} & \frac{4}{6} & -\frac{2}{6} \\ \frac{1}{6} & -\frac{2}{6} & \frac{1}{6} \end{pmatrix}$$

Method 8.4.4. . Given an orthogonal basis $\mathcal{B} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_n)$ for $\mathbb{R}^n$ and scalars $\alpha_1, \alpha_2, \dots, \alpha_n$ find a symmetric matrix A such that $\boldsymbol{w}_i$ is an eigenvector of A with eigenvalue α_i.

If necessary, normalize the basis vectors in $\mathcal{B}$ to obtain an orthonormal basis $\mathcal{O} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$. Set

$$A = \alpha_1 \boldsymbol{x}_1 \boldsymbol{x}_1^{Tr} + \alpha_2 \boldsymbol{x}_2 \boldsymbol{x}_2^{Tr} + \cdots + \alpha_n \boldsymbol{x}_n \boldsymbol{x}_n^{Tr}.$$

A is the desired matrix.

Example 8.4.13. Find a symmetric matrix A with eigenvalues 5 and -3 with corresponding eigenvectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

We normalize the vectors to obtain the orthonormal basis

$$\mathcal{O} = \left(\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right)$$

$$A = 5 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}^{Tr} - 3 \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}^{Tr} =$$

$$5 \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} - 3 \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}$$

Example 8.4.14. Find a symmetric matrix A with eigenvalues 4, 3 and 1 and corresponding eigenvectors $\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$, $\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$, respectively.

We normalize the vectors to obtain the orthonormal basis

$$\mathcal{O} = \left(\begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \right)$$

$$A = 4 \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}^{Tr} + 3 \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}^{Tr} + \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}^{Tr} =$$

$$4\begin{pmatrix} \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} + 3\begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} \frac{4}{6} & \frac{2}{6} & \frac{2}{6} \\ \frac{2}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{2}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} =$$

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

Exercises

In 1 - 16, for each **symmetric matrix**, A, find an **orthogonal matrix** Q and a **diagonal matrix** D such that $Q^{Tr}AQ = D$. See **Method** (8.4.2).

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

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

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

4. $A = \begin{pmatrix} -3 & 3 \\ 3 & 5 \end{pmatrix}$

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

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

7. $A = \begin{pmatrix} 2 & 12 \\ 12 & -5 \end{pmatrix}$

8. $A = \begin{pmatrix} 10 & 12 \\ 12 & 17 \end{pmatrix}$

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

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

11. $A = \begin{pmatrix} 5 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 3 \end{pmatrix}$

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

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

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

15. $A = \begin{pmatrix} 5 & -1 & -3 \\ -1 & 5 & -3 \\ -3 & -3 & -3 \end{pmatrix}$

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

For each of the matrices in exercises 17-24 obtain a **spectral decomposition**. See **Method** (8.4.3).

17. The matrix of exercises 2.

18. The matrix of exercise 4.

19. The matrix of exercise 6.

20. The matrix of exercise 8.

21. The matrix of exercises 10.

22. The matrix of exercise 12.

23. The matrix of exercise 14.

24. The matrix of exercise 16.

In exercises 25 - 30 find a **symmetric matrix** A with the given **eigenvalues** and corresponding **eigenvectors**. See **Method** (8.4.4).

25. **Eigenvalues** 1, 5 and 7 with **eigenvectors** $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, respectively.

26. **Eigenvalues** 1, 1 and 7 with **eigenvectors** $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, respectively.

27. **Eigenvalues** 3, -1 and -3 with **eigenvectors** $\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$, respectively.

28. **Eigenvalues** 2, 5 and -7 with **eigenvectors** $\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}$, respectively.

29. **Eigenvalues** 6, -3 and -3 with **eigenvectors** $\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}$, respectively.

30. **Eigenvalues** 5, -4 and -4 with **eigenvectors** $\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix}$, respectively.

In exercises 31 - 36 answer true or false and give an explanation.

31. If $A = Q^T DQ$ where Q is an **orthogonal matrix** and D is a **diagonal matrix** then A is **symmetric**.

32. If A is a **symmetric** then A is **orthogonally diagonalizable**.

33. If A is **symmetric** then every **eigenvalue** of A has **algebraic multiplicity** one.

34. If A is a **symmetric matrix** and $\boldsymbol{u}, \boldsymbol{v} \in null(A)$ then $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

35. If A is a **symmetric matrix**, $\boldsymbol{u} \in null(A)$ and $A\boldsymbol{v} = -\boldsymbol{v}$ then $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

36. If A is a **symmetric matrix** and α is an **eigenvalue** , then the **algebraic multiplicity** of α is equal to the dimension of the $\alpha-$**eigenspace** of A.

Challenge Exercises (Problems)

All the matrices in these challenge exercises are assumed to be real matrices.

1. Assume A is a 3×3 **symmetric matrix** with integer entries. Assume that $a \neq 0$ is an **eigenvalue** with **eigenspace** $Span\left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}\right)$ and $b \neq 0$ is an **eigenvalue** with **eigenspace** $Span\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}\right)$. Find A such that $|ab|$ is as small as possible.

2. Assume that A and B are **orthogonally diagonalizable** $n \times n$ matrices and $AB = BA$. Prove that there exists an **orthonormal basis** for $\mathbb{R}^n$ consisting of vectors which are **eigenvectors** for both A and B.

3. Assume that A and B are **orthogonally diagonalizable** $n \times n$ matrices and $AB = BA$. Prove that AB is **orthogonally diagonalizable**. See **Theorem** (8.4.4).

4. Assume that A is **orthogonally diagonalizable** and **invertible**. Prove that A^{-1} is **orthogonally diagonalizable**. See **Theorem** (8.4.4).

5. Assume that A is an $n \times n$ **orthogonally diagonalizable** matrix. Let $f(x) = c_d x^d + c_{d-1}x^{d-1} + \cdots + c_1 x + c_0$ be a polynomial. Recall that by $f(A)$ we mean $c_d A^d + c_{d-1}A^{d-1} + \cdots + c_1 A + c_0 I_n$. Prove that $f(A)$ is **orthogonally diagonalizable**. See **Theorem** (8.4.4)).

6. a) Assume that A is an **invertible** $n \times n$ matrix and set $B = A^{Tr}A$. Prove that B is a **symmetric matrix** and that all the **eigenvalues** of B are positive.

b) Assume that B is a **symmetric matrix** and that all the **eigenvalues** of B are positive. Prove that there is a **symmetric** matrix S such that $B = S^2$.

c) Assume that B is a **symmetric** and that all the **eigenvalues** of B are positive. Prove that there is an **invertible matrix** A such that $B = A^{Tr}A$.

7. Let A be an $m \times m$ **symmetric matrix** . Assume A has zero as an eigenvalue z times, p positive **eigenvalues** (counting multiplicity) and n negative **eigenvalues** (counting multiplicity).

a) Let $\mathcal{N}$ consist of all **subspaces** N of $\mathbb{R}^n$ such that for every non-zero vector $\boldsymbol{n} \in N, \boldsymbol{n}^{Tr}A\boldsymbol{n} < 0$. Amongst all **subspaces** in $\mathcal{N}$ choose one, N', with the largest dimension. Prove that the **dimension** of N' is n.

b) Let $\mathcal{P}$ consist of all **subspaces** P of $\mathbb{R}^n$ such that for every non-zero vector $\boldsymbol{p} \in P, \boldsymbol{p}^{Tr}A\boldsymbol{p} > 0$. Amongst all **subspaces** in $\mathcal{P}$ choose one, P', with the largest dimension. Prove that the **dimension** of P' is p.

8. A **linear transformation** $T : \mathbb{R}^n \to \mathbb{R}^n$ is called ***self adjoint*** if for every pair of vectors $\boldsymbol{x}, \boldsymbol{y}, T(\boldsymbol{x}) \cdot \boldsymbol{y} = \boldsymbol{x} \cdot T(\boldsymbol{y})$. Prove that T is self adjoint if and only if the **standard matrix** of T is **symmetric**.

Quiz Solutions

1. $\left(\begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}, \begin{pmatrix} \frac{4}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \end{pmatrix}\right)$. Not right, see the Gram-Schmidt process, **Method** (8.2.1).

2. $\left(\begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{pmatrix}\right)$. Not right, see the Gram-Schmidt process, **Method** (8.2.1).

3. The sequence $\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right)$ is a **basis** consisting of **eigenvectors** of A with respective **eigenvalues** 1 and 5. Not right, see **Method** (7.1.2).

4. This matrix has **eigenvalues** 3 with multiplicity one and 6 with multiplicity two. The vector $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is an **eigenvector** with **eigenvalue** 3. The sequence $\left(\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right)$ is a **basis** for E_6, the 6-**eigenspace**. Thus, $\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right)$ is a **basis** for $\mathbb{R}^3$ consisting of **eigenvectors** of A. Not right, see **Method** (7.1.2).

8.5. Quadratic Forms, Conic Sections and Quadratic Surfaces

In this section we introduce the notion of a quadratic polynomial in n variables, a special case of which is a quadratic form. We associate to each quadratic form a **symmetric matrix**. We use **orthogonal diagonalization** to classify quadratic forms. We go on to show that the zero set of a non-degenerate plane quadratic polynomial is a hyperbola, ellipse (circle) or parabola and the zero set of a non-degenerate quadratic polynomial in three space is an ellipsoid (a sphere is a special case), hyperboloid of one or two sheets, an elliptic paraboloid, or a hyperbolic paraboloid.

Am I Ready for This Material

Readiness Quiz

New Concepts

Theory (Why It Works)

What You Can Now Do

Method (How To Do It)

Exercises

Challenge Exercises (Problems)

Am I Ready for This Material

The following are essential to an understanding of the material introduced in this section:

symmetric matrix

similar matrices

eigenvector

eigenvalue

dot product

orthogonal or perpendicular vectors

orthogonal basis of a subspace of $\mathbb{R}^n$

orthonormal basis of a subspace of $\mathbb{R}^n$

orthogonal matrix

orthogonally diagonalizable matrix

Quiz

1. Let $A = \begin{pmatrix} -1 & 2 \\ 2 & 2 \end{pmatrix}$. Find an **orthogonal matrix** Q such that $Q^{Tr}AQ$ is a diagonal matrix.

2. Let $A = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$. Find an **orthogonal matrix** Q such that $Q^{Tr}AQ$ is a diagonal matrix.

3. Let $A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$. Find an **orthogonal matrix** Q such that $Q^{Tr}AQ$ is a diagonal matrix.

4. Let $A = \begin{pmatrix} 3 & 0 & -1 \\ 0 & 3 & -1 \\ -1 & -1 & 2 \end{pmatrix}$. Find an **orthogonal matrix** Q such that $Q^{Tr}AQ$ is a diagonal matrix.

Quiz Solutions

New Concepts

In this section we introduce several new definitions. These are:

quadratic polynomial

quadratic form

matrix of a quadratic form

conic section

quadratic surface

quadric

positive definite quadratic form

positive semi-definite quadratic form

negative definite quadratic form

negative semi-definite quadratic form

indefinite quadratic form

positive definite matrix

positive semi-definite matrix

negative definite matrix

negative semi-definite matrix

indefinite matrix

Theory (Why It Works)

We begin with several definitions.

Definition 8.12. Let $X_1, \dots, X_n$ be variables (indeterminates). By a ***linear form*** on $X_1, \dots, X_n$ we mean a non-trivial **linear combination** of $X_1, \dots, X_n$. By a ***polynomial*** of degree one we mean the sum of a linear form and a scalar.

By a ***quadratic form*** we shall mean a non-trivial **linear combination** of the monomials $X_i^2, 1 \leq i \leq n$ and X_iX_j where $1 \leq i < j \leq n$.

Finally, by a ***quadratic polynomial*** we shall mean a quadratic form or a sum of a quadratic form, a linear form, and/or a scalar. When $\hat{p} = \sum_{i=1}^n a_iX_i^2 + \sum_{i<j} a_{ij}X_iX_j + \sum_{i=1}^n b_iX_i + c$ then the quadratic form $\sum_{i=1}^n a_iX_i^2 + \sum_{i<j} a_{ij}X_iX_j$ is referred to as the ***quadratic part*** of $\hat{p}$.

Example 8.5.1. Let $\hat{q} = X_1^2 + 3X_2^2 - 4X_3^2 + 3X_1X_2 - 2X_1X_3 + 7X_2X_3$. Then $\hat{q}$ is a quadratic form in three variables.

The polynomial $\hat{p} = X_1^2 - X_2X_3 + 4X_2$ is a quadratic polynomial in three variables, but not a quadratic form. The quadratic part of $\hat{p}$ is $X_1^2 - X_2X_3$. On the other hand, $\hat{l} = 2X_1 - 5X_2 + 7X_3$ is not a quadratic form.

Example 8.5.2. Let $\hat{q} = 2X_1^2 - 3X_2^2 + X_1X_2 - 4X_1X_3 + 5X_2X_3$, a quadratic form in three variables. Define a matrix A as follows: The diagonal entries are the coefficients of X_1^2, X_2^2, X_3^2, respectively. The (1,2) and (2,1) entries are each one half the coefficient of X_1X_2; the (1,3) and (3,1) entries are each one half the coefficient of X_1X_3; and the (2,3) and (3,2) entries are each one half the coefficient of X_2X_3. Thus,

$$A = \begin{pmatrix} 2 & \frac{1}{2} & -2 \\ \frac{1}{2} & -3 & \frac{5}{2} \\ -2 & \frac{5}{2} & 0 \end{pmatrix}$$

As a consequence of how we have defined the matrix A it is necessarily **symmetric**.

Also observe that if we set $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ then $\hat{q} = \boldsymbol{X}^{Tr}A\boldsymbol{X}$.

There is nothing special about the quadratic form of Example (8.5.2), we can define a **symmetric matrix** in a similar way for any quadratic form:

Definition 8.13. Let $\hat{q} = \sum_{i=1}^{n} b_i X_i^2 + \sum_{1\le i<j\le n} b_{ij}X_iX_j$ be a quadratic form in the n variables $X_1, \dots, X_n$. Define the matrix A as follows: the (i,i) entry of A, a_{ii}, is b_i; for $1 \le i < j \le n$ the (i,j) and (j,i) entries are equal to one half of $b_{ij}, a_{ij} = a_{ji} = \frac{b_{ij}}{2}$. The matrix A is **symmetric** and called the ***matrix*** of the quadratic form $\hat{q}$.

Remark 8.5. Alternatively, we could define a **quadratic form** to be an expression of the form $(X_1 \ \dots \ X_n)A\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ where A is a **symmetric matrix**.

Remark 8.6. Let $\hat{p}$ be a **quadratic polynomial** in the variables $X_1, \dots, X_n$ with **quadratic part** $\hat{q}$ and let A be the **matrix** of $\hat{q}$. Then there is a row matrix B and scalar C such that $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + C$ where $\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$.

Definition 8.14. A **quadratic form** $\hat{q}$ is said to have ***standard form*** if its matrix is diagonal.

Example 8.5.3. Let

$$\hat{p} = -2X_1^2 + 3X_2^2 - 3X_3^2 + X_1X_2 + 4X_1X_3 - 5X_2X_3 + 6X_1 - X_2 + X_3 - 6.$$

Set $A = \begin{pmatrix} -2 & \frac{1}{2} & 2 \\ \frac{1}{2} & 3 & -\frac{5}{2} \\ 2 & -\frac{5}{2} & -3 \end{pmatrix}, B = \begin{pmatrix} 6 \\ -1 \\ 1 \end{pmatrix}^{Tr} = (6 \ -1 \ 1), C = -6$. Then $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + C$.

The reason we have introduced these concepts is in order to define a quadratic curve (in $\mathbb{R}^2$) and quadratic surface (in $\mathbb{R}^3$.)

Definition 8.15. By a ***conic section*** or a plane ***quadratic curve*** we mean the collection of vectors $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ such $\hat{p}(\boldsymbol{x}) = 0$ for a fixed **quadratic polynomial** $\hat{p}$ of X_1, X_2.

Definition 8.16. By a ***quadratic surface*** in $\mathbb{R}^3$ we mean the collection of vectors $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ such that $\hat{p}(\boldsymbol{x}) = 0$ for a fixed **quadratic polynomial** $\hat{p}$ of X_1, X_2, X_3.

When $\hat{p}$ is a **quadratic form** we usually refer to the quadratic surface defined by $\hat{p}$ as a ***quadric***.

More generally, by a ***quadratic hypersurface*** in $\mathbb{R}^n$ we mean the collection of vectors $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ such that $\hat{p}(\boldsymbol{x}) = 0$ for a fixed **quadratic polynomial** $\hat{p}$ of the variables $X_1, \dots, X_n$. When $\hat{p}$ is a **quadratic form** the hypersurface defined by $\hat{p}$ is called a ***quadric***.

For a **quadratic polynomial** $\hat{p}$ of n variables we denote by $Z(\hat{p})$ the collection

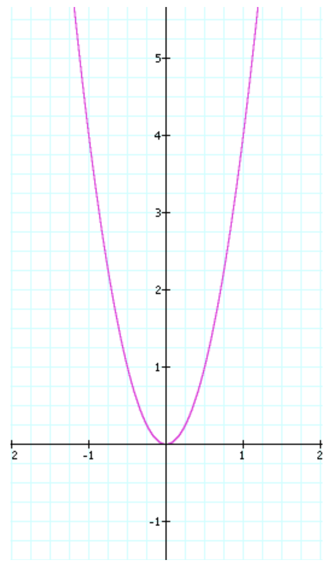
$$\{\boldsymbol{x} \in \mathbb{R}^n : \hat{p}(\boldsymbol{x}) = 0\}.$$

Remark 8.7. If $\hat{p}$ is a **quadratic polynomial** and $c \neq 0$ is a scalar then $Z(\hat{p})$ and $Z(c\hat{p})$ are identical.

In the next several examples we illustrate what kinds of planar graphs arise as a conic section for a **quadratic polynomial** $\hat{p}$ when the **quadratic part** is in **standard form**, that is, there is no cross product term, X_1X_2.

Example 8.5.4. Let $\hat{p}(\boldsymbol{x}) = aX_1^2 - X_2$. Then $Z(\hat{p})$ is a **parabola** with a vertical axis of symmetry. It opens up if $a > 0$ and down if $a < 0$. This is shown in Figure (8.5.1).

Figure 8.5.1: Parabola with vertical axis of symmetric



If $\hat{p} = aX_2^2 - X_1$ then $Z(\hat{p})$ is again a **parabola** , now with a horizontal axis of symmetry.

Example 8.5.5. Let $\hat{p} = \frac{X_1^2}{a^2} + \frac{X_2^2}{b^2} - 1$ with $a, b > 0$. $Z(\hat{p})$ is an **ellipse** with axes of length $2a$ and $2b$ (when $a = b$ it is a circle with radius a). This is shown in Figure(8.5.2).

Example 8.5.6. Let $\hat{p} = \frac{X_1^2}{a^2} - \frac{X_2^2}{b^2} - 1$ or $\frac{X_2^2}{b^2} - \frac{X_1^2}{a^2} - 1$ with $a, b > 0$. Then $Z(\hat{p})$ is a **hyperbola**. An example is shown in Figure (8.5.3).

We will shortly show that if $\hat{p}$ is a **quadratic polynomial** in two variables then $Z(\hat{p})$ is a parabola, ellipse (circle), hyperbola, a pair of intersecting lines, a point or the empty set.

Figure 8.5.2: Ellipse with horizontal major axis

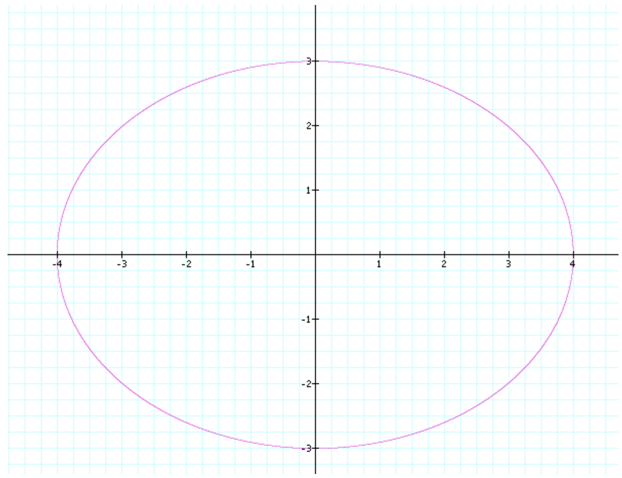
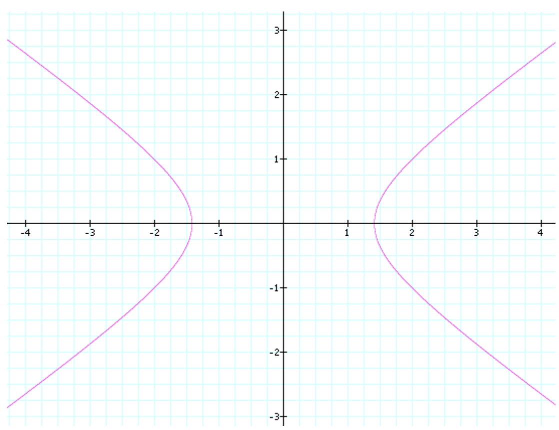


Figure 8.5.3: Graph of a hyperbola

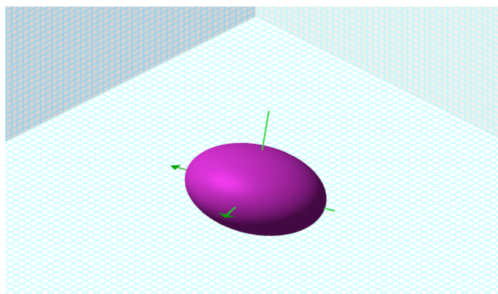


In following examples we consider **quadric surfaces** which are defined as the zeros of a **quadratic polynomial** in three variables when there are no cross product terms (X_1X_2, X_1X_3, X_2X_3).

Example 8.5.7. Let $\hat{p} = \frac{X_1^2}{a^2} + \frac{X_2^2}{b^2} + \frac{X_3^2}{c^2} - 1$ with $a, b, c > 0$. $Z(\hat{p})$ is an **ellipsoid**.

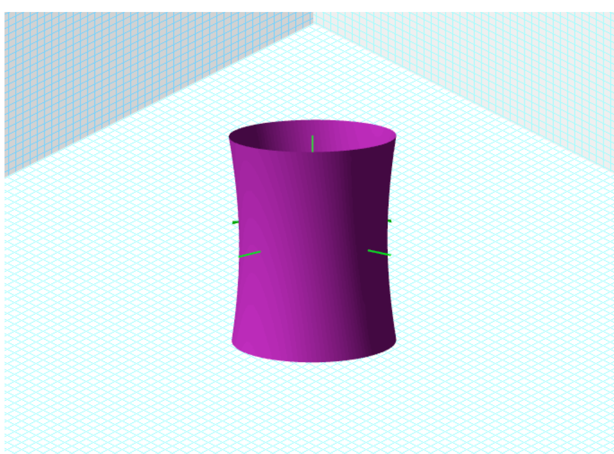
(When $a = b = c$ is it a sphere.) See Figure (8.5.4) for an example.

Figure 8.5.4: Graph of a ellipsoid



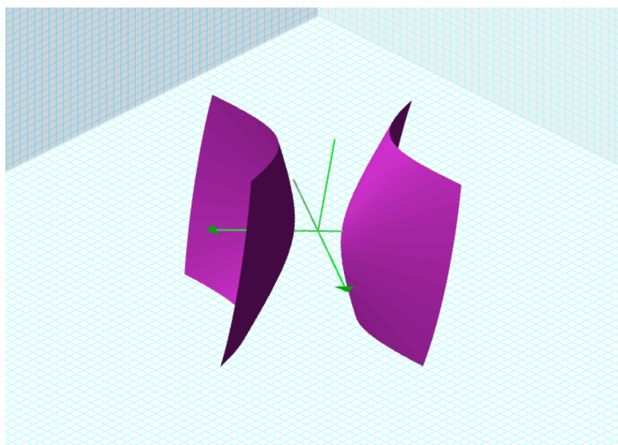
Example 8.5.8. Now, let $\hat{p} = \frac{X_1^2}{a^2} + \frac{X_2^2}{b^2} - \frac{X_3^2}{c^2} - 1$ with $a, b, c > 0$. In this case $Z(\hat{p})$ is called a **hyperboloid of one sheet**. An example is shown in Figure (8.5.5).

Figure 8.5.5: Graph of a hyperboloid of one sheet



Example 8.5.9. Set $\hat{p} = \frac{X_1^2}{a^2} + \frac{X_2^2}{b^2} - \frac{X_3^2}{c^2} + 1$ with $a, b, c > 0$. In this case $Z(\hat{p})$ is called a **hyperboloid of two sheets**. See Figure (8.5.6) for an example of this graph.

Figure 8.5.6: Graph of a hyperboloid of two sheets



Example 8.5.10. If $\hat{p} = \frac{X_1^2}{a^2} + \frac{X_2^2}{b^2} - X_3$ with $a, b > 0$ then $Z(\hat{p})$ is an **elliptic paraboloid**. A graph is shown in Figure (8.5.7).

Example 8.5.11. When $\hat{p} = \frac{X_1^2}{a^2} - \frac{X_2^2}{b^2} - X_3$, $Z(\hat{p})$ is a **hyperbolic paraboloid**. See Figure (8.5.8) for an example of such a graph.

Figure 8.5.7: Graph of a elliptic paraboloid

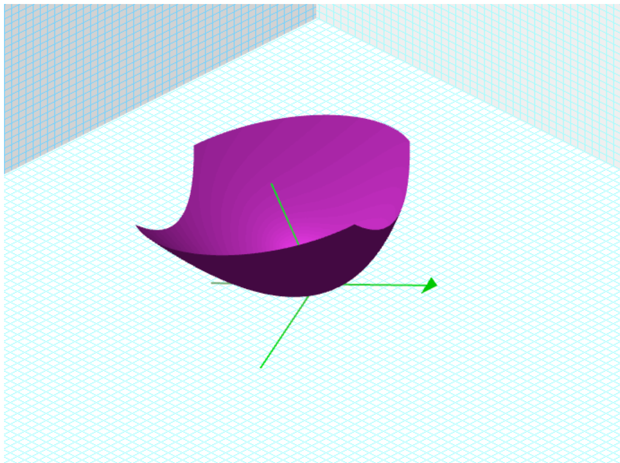
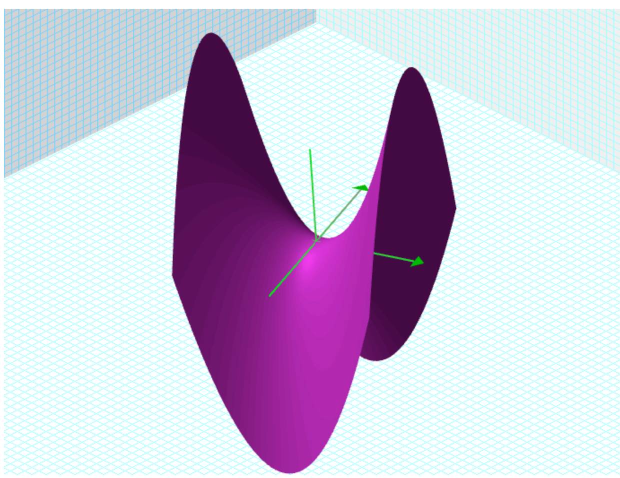


Figure 8.5.8: Graph of a hyperbolic paraboloid



We shall show that if $\hat{p}$ is a **quadratic polynomial** in three variables (not reducible to a polynomial in two variables) then $Z(\hat{p})$ is either an ellipsoid (sphere), hyperboloid of one sheet, hyperboloid of two sheets, an elliptic cone, an elliptic paraboloid or a hyperbolic paraboloid. The next example illustrates how we do this when $\hat{p}$ has no cross product terms.

Example 8.5.12. Let $\hat{p} = X_1^2 - 3X_2^2 + 4X_3^2 - 2X_1 + 3X_2 - 6X_3$. Describe $Z(\hat{p})$.

We will make a change of variable via translation that will eliminate the linear terms and thereby express $\hat{p}$ as a **quadratic form** (without cross product terms) plus a scalar.

Thus, set

$$X_1' = X_1 - 1, X_2' = X_2 - \frac{1}{2}, X_3' = X_3 - \frac{3}{4}$$

Then

$$(X_1')^2 - 3(X_2')^2 + 4(X_3')^2 = X_1^2 - 2X_1 + 1 - 3(X_2^2 - X_2 + \frac{1}{4}) + 4(X_3^2 - \frac{3}{2}X_3 + \frac{9}{16}) =$$

$$X_1^2 - 3X_2^2 + 4X_3^2 - 2X_1 + 3X_2 - 6X_3 + (1 - \frac{3}{4} + \frac{9}{4}) = \hat{p} + \frac{5}{2}$$

Therefore $\hat{p} = (X_1')^2 - 3(X_3')^2 + 4(X_3')^2 - \frac{5}{2}$. The **zero set** of $(X_1')^2 - 3(X_2')^2 + 4(X_3')^2 - \frac{5}{2}$ is a hyperboloid of one sheet in $X_1'X_2'X_3'$ space.

We obtain the graph of $\hat{p}$ in $X_1X_2X_3$ space by translating the graph to the point $\begin{pmatrix} 1 \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix}$.

What we did in **Example** (8.5.11) can be done in general. We state and prove this for the special case where the **matrix** of the **quadratic part** of the **quadratic polynomial** $\hat{p}$ is **diagonal** and **invertible**.

Theorem 8.5.1. *Let $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + C$ be a **quadratic polynomial** and assume the matrix A is **diagonal** and **invertible**. Then there is a unique vector $\boldsymbol{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$ and scalar C' such that if $\boldsymbol{X}' = \boldsymbol{X} - \boldsymbol{d}$ then*

$$\hat{p}(\boldsymbol{x}) = (\boldsymbol{X}')^{Tr}A(\boldsymbol{X}') + C' \tag{8.43}$$

Proof. Let $A = diag(a_1, a_2, \ldots, a_n)$ and $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}^{Tr}$. Note that since A is invertible it follows that for each $i, a_i \neq 0$.

Set $d_i = -\frac{b_i}{2a_i}, \boldsymbol{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$ and $C' = C - \sum_{i=1}^n \frac{b_i^2}{4a_i}$. Then

$$(\boldsymbol{X}')^{Tr}A(\boldsymbol{X}') + C' = \sum_{i=1}^n a_i(X_i - d_i)^2 + (C - \sum_{i=1}^n \frac{b_i^2}{4a_i}) =$$

$$\sum_{i=1}^n a_i(X_i + \frac{b_i}{2a_i})^2 + (C - \sum_{i=1}^n \frac{b_i^2}{4a_i}) =$$

$$\sum_{i=1}^n a_i(X_i^2 + \frac{b_i}{a_i}X_i + \frac{b_i^2}{4a_i^2}) + (C - \sum_{i=1}^n \frac{b_i^2}{4a_i}) =$$

$$\sum_{i=1}^n (a_iX_i^2 + b_iX_i + \frac{b_i^2}{4a_i}) + (C - \sum_{i=1}^n \frac{b_i^2}{4a_i}) =$$

$$\sum_{i=1}^n a_iX_i^2 + \sum_{i=1}^n b_iX_i + (\sum_{i=1}^n \frac{b_i^2}{4a_i} + C - \sum_{i=1}^n \frac{b_i^2}{4a_i}) = \hat{p}.$$

This shows that $\boldsymbol{d}, C'$ satisfy the requirements of the theorem. We must show that they are unique. Suppose then that $\boldsymbol{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$ and C^* satisfy (8.43). Then

$$\sum_{i=1}^n a_i(X_i - d_i)^2 + C^* = \sum_{i=1}^n a_iX_i^2 + \sum_{i=1}^n b_iX_i + C \tag{8.44}$$

It follows from (8.44) that

$$\sum_{i=1}^n a_i(X_i^2 - 2d_iX_i + d_i^2) + C^* = \sum_{i=1}^n a_iX_i^2 + \sum_{i=1}^n b_iX_i + C \tag{8.45}$$

Then

$$\sum_{i=1}^n a_iX_i^2 - \sum_{i=1}^n 2a_id_iX_i + \sum_{i=1}^n a_id_i^2 + C^* = \sum_{i=1}^n a_iX_i^2 + \sum_{i=1}^n b_iX_i + C \tag{8.46}$$

It is therefore the case that $-2a_id_i = b_i$ from which we conclude that $d_i = -\frac{b_i}{2a_i}$.

We also conclude that

$$\sum_{i=1}^n a_id_i^2 + C^* = C \tag{8.47}$$

Substituting $d_i = -\frac{b_i}{2a_i}$ into (8.47) we get that $C^* = C - \sum_{i=1}^{n} \frac{b_i^2}{4a_i}$ as claimed and completes the proof of the theorem $\square$

Remark 8.8. When $A = diag(a_1, \ldots, a_n)$, is **diagonal** but some of the a_i are zero we can eliminate the linear terms for all a_i which are non-zero.

Example 8.5.13. Let $\hat{p} = X_1^2 - 2X_2^2 + 2X_1 - 8X_2 + 3X_3$. Eliminate the linear terms $2X_1$ and $-8X_2$.

Set $X_1' = X_1 + 1, X_2' = X_2 + 2$. Then

$$(X_1')^2 - 2(X_2')^2 = (X_1+1)^2 - 2(X_2+2)^2 = (X_1^2 + 2X_1 + 1) - 2(X_2^2 + 4X_2 + 4) =$$

$$X_1^2 - 2X_2^2 + 2X_1 - 8X_2 - 7 = \hat{p} - 3X_3 - 7$$

Thus,

$$\hat{p} = (X_1')^2 - 2(X_2')^2 + 3X_3 + 7.$$

If we further set $X_3' = X_3 + \frac{7}{3}$ then

$$\hat{p} = (X_1')^2 - 2(X_2')^2 + 3X_3'.$$

The zero set of $\hat{p}' = (X_1')^2 - 2(X_2')^2 + 3X_3'$ in $X_1'X_2'X_3'$ space is a hyperbolic paraboloid. We obtain the graph for $Z(\hat{p})$ by translating the graph $Z(\hat{p}')$ to the point $\begin{pmatrix} -1 \\ -2 \\ -\frac{7}{3} \end{pmatrix}$.

Let $\hat{p}$ be a **quadratic polynomial** with **quadratic part** $\hat{q}$ and assume that the **matrix** of $\hat{q}$ is diagonal. We can see from **Theorem** (8.5.1) and **Example** (8.5.12) that it is fairly straightforward to describe the **quadratic surface** defined by $\hat{p}$ by using a change of variable obtained by translating the origin. Fortunately, it is also possible, using an **orthogonal change of variable**, to diagonalize any **quadratic form**. First we define what we mean by an ***orthogonal change of variable***.

Definition 8.17. Let $\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ be a variable vector in $\mathbb{R}^n$. By a ***linear change of variable*** we mean a variable $\boldsymbol{Y} = Q^{-1}\boldsymbol{X}$ where Q is an **invertible matrix**. When Q is an **orthogonal matrix** and $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$, we say that $\boldsymbol{Y}$ is obtained by an ***orthogonal change of variable***.

Definition 8.18. Let $\hat{q}$ and $\hat{q}'$ be **quadratic forms** in n variables with respective **matrices** A and A'. If there exists an **orthogonal matrix** Q such that $A' = Q^{Tr}AQ$ then we say the **quadratic forms** $\hat{q}$ and $\hat{q}'$ are ***orthogonally equivalent***.

Assume that $\boldsymbol{Y}$ is obtained from $\boldsymbol{X}$ by a linear change of variable where $\boldsymbol{X} = Q\boldsymbol{Y}$ and that $\hat{q}$ is a **quadratic form** with matrix A so that $\hat{q} = \boldsymbol{X}^{Tr}A\boldsymbol{X}$. Then

$$\hat{q} = (Q\boldsymbol{Y})^{Tr}A(Q\boldsymbol{Y}) = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} \tag{8.48}$$

Let $\hat{q}'$ be the **quadratic form** with matrix $Q^{Tr}AQ$. Then $\hat{q}(\boldsymbol{X}) = \hat{q}'(\boldsymbol{Y})$.

Let A be a **symmetric matrix**. Recall by the **Spectral Theorem** there is an **orthogonal matrix** Q and a **diagonal matrix** D such that $Q^{Tr}AQ = D$. For such an **orthogonal matrix** Q, the resulting form $\hat{q}'$ has no cross product terms, that is, $\hat{q}'$ is in **standard form**. We summarize this in a result usually referred to as the **Principal Axis Theorem**.

Theorem 8.5.2. *Let $\hat{q}$ be a **quadratic form** with **matrix** A. Then there is an orthogonal change of variable $\boldsymbol{X} = Q\boldsymbol{Y}$ that transforms $\hat{q}$ into a **quadratic form** $\hat{q}'$ with a diagonal **matrix** $Q^{Tr}AQ$.*

Example 8.5.14. Let $\hat{q} = X_1^2 - 2X_2^2 + 4X_1X_2$. Find an orthogonal change of variable which transforms $\hat{q}$ into a **quadratic form** without cross product terms.

The **matrix** of $\hat{q}$ is $A = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}$. The **characteristic polynomial** of A is

$$\chi_A(\lambda) = \lambda^2 + \lambda - 6 = (\lambda + 3)(\lambda - 2)$$

Therefore the **eigenvalues** are -3 and 2. Corresponding **eigenvectors** are: $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Notice that these are **orthogonal**. We **normalize** to obtain an **orthonormal**

basis: $\left(\begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}\right)$. Set $Q = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$ and $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$ so that $\boldsymbol{X} = Q\boldsymbol{Y}$.

Then

$$\hat{q}(\boldsymbol{X}) = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} = (Y_1\ Y_2)\begin{pmatrix} -3 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = -3Y_1^2 + 2Y_2^2 = \hat{q}'(\boldsymbol{Y}).$$

Our ultimate goal is to classify the **conic sections** (quadratic curves in $\mathbb{R}^2$) and the **quadratic surfaces** in $\mathbb{R}^3$. Before preceding we discuss a way of categorizing **quadratic forms**.

Classifying quadratic forms

Let $\hat{q}$ be a **quadratic form** in n variables $X_1, \dots, X_n$. Then $\hat{q}$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$. We categorize the form $\hat{q}$ by the values it takes when evaluated at non-zero vectors $\boldsymbol{x} : \{\hat{q}(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0}_n\}$.

Definition 8.19. A **quadratic form** $\hat{q}$ is

i. ***positive definite*** if $\hat{q}(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}_n$ in $\mathbb{R}^n$;

ii. ***negative definite*** if $\hat{q}(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}_n$ in $\mathbb{R}^n$;

iii. ***indefinite*** if $\hat{q}$ takes both positive and negative values;

iv. ***positive semi-definite*** if $\hat{q}(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$; and

v. ***negative semi-definite*** if $\hat{q}(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Definition 8.20. Let A be a **symmetric matrix**. We will say that A is ***positive definite*** if and only if the **quadratic form** $\hat{q}$ with matrix A is positive definite. In a similar way we define A to be negative definite, positive semi-definite, negative semi-definite or indefinite if and only if the **quadratic form** $\hat{q}$ with matrix A is negative definite, positive semi-definite, negative semi-definite or indefinite, respectively.

For a **quadratic form** $\hat{q}$ with **matrix** A we can determine the type of $\hat{q}$ entirely from the **eigenvalues** of A :

Theorem 8.5.3. *Let $\hat{q}$ be a* **quadratic form** *with* **matrix** *A. Then $\hat{q}$ is*

i. positive semidefinite if and only if the **eigenvalues** *of A are all non-negative;*

ii. positive definite if and only if the **eigenvalues** *of A are all positive;*

iii. negative semidefinite if and only if the **eigenvalues** *of A are all non-positive;*

iv. negative definite if and only if the **eigenvalues** *of A are all negative; and*

v. indefinite if and only if A has both positive and negative **eigenvalues** *.*

Proof. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the eigenvalues of the symmetric matrix A and let $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \ldots\ \boldsymbol{v}_n)$ be an orthonormal matrix such that $\boldsymbol{v}_i$ is an **eigenvector** of A with **eigenvalue** α_i so that $Q^{Tr}AQ = diag(\alpha_1, \alpha_2, \ldots, \alpha_n) = D$. We saw in the **Principal Axis Theorem** that if $\boldsymbol{X} = Q\boldsymbol{Y}$ then $\hat{q}(\boldsymbol{X}) = \hat{q}'(\boldsymbol{Y})$ where $\hat{q}'$ is the **quadratic form** with **matrix** D :

$$\hat{q}(\boldsymbol{X}) = \hat{q}'(\boldsymbol{Y}) = \alpha_1 Y_1^2 + \cdots + \alpha_n Y_n^2.$$

Since Q is invertible, $\boldsymbol{y} \neq \boldsymbol{0}_n$ if and only if $\boldsymbol{x} = Q\boldsymbol{y} \neq \boldsymbol{0}_n$. Therefore

$$\{\hat{q}(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0}_n\} = \{\hat{q}'(\boldsymbol{y}) : \boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{y} \neq \boldsymbol{0}_n\}.$$

It then follows that $\hat{q}$ and $\hat{q}'$ are the same type so we need only establish the theorem for $\hat{q}'$. The proofs of all five conditions are quite similar and so we illustrate with the second and fifth.

ii. Suppose all the α_i are positive and $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \neq \boldsymbol{0}_n$. Each $y_i^2 \geq 0$ and since $\alpha_i > 0$ it follows that $\alpha_i y_i^2 \geq 0$. Since the sum of non-negative numbers is non-negative it follows that

$$\hat{q}'(\boldsymbol{y}) = \alpha_1 y_1^2 + \alpha_2 y_2^2 + \cdots + \alpha_n y_n^2 \geq 0$$

On the other hand, since $\boldsymbol{y} \neq \boldsymbol{0}_n$ for some $i, y_i \neq 0$. Then $\alpha_i y_i^2 > 0$ and $\hat{q}'(\boldsymbol{y}) > 0$.

On the other hand, suppose that $\hat{q}'$ is positive definite but to the contrary that for some $i, \alpha_i \leq 0$. Then $\hat{q}'(\boldsymbol{e}_i) = \alpha_i \leq 0$ (where $\boldsymbol{e}_i$ is the i^{th} standard basis vector), a contradiction. Thus, all $\alpha_i > 0$.

v. Suppose there are both positive and negative eigenvalues, say $\alpha_i > 0$ and $\alpha_j < 0$. Then $\hat{q}'(\boldsymbol{e}_i) = \alpha_i > 0$ and $\hat{q}'(\boldsymbol{e}_j) = \alpha_j < 0$ and $\hat{q}'$ is indefinite.

Conversely, assume that $\hat{q}'$ is indefinite. If all the α_i are non-negative then the argument used in i. implies that $\hat{q}'(\boldsymbol{y}) \geq 0$ for all $\boldsymbol{y}$ and therefore not indefinite. Similarly, if all the α_i are non-positive then $\hat{q}'(\boldsymbol{y}) \leq 0$ for all $\boldsymbol{y}$ and consequently not indefinite. Thus, there must be both positive and negative **eigenvalues** . □

Example 8.5.15. Let $\hat{q}$ be the **quadratic form** with **matrix** $A = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & 2 \\ -2 & 2 & 3 \end{pmatrix}$.

Determine the type of $\hat{q}$.

The **characteristic polynomial** of A is $\chi_A(\lambda) =$

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

$$(2-\lambda)(2-\lambda)(3-\lambda) + (1)(2)(-2) + (-2)(1)(2)$$

$$-(2-\lambda)(2)(2) - (1)(1)(3-\lambda) - (-2)(2-\lambda)(-2) =$$

$$-\lambda^3 + 7\lambda^2 - 16\lambda + 12 - 8 + 4\lambda - 3 + \lambda - 8 + 4\lambda$$

$$-\lambda^3 + 7\lambda^2 - 7\lambda - 15 = -(\lambda^3 - 7\lambda^2 + 7\lambda + 15) = -(\lambda+1)(\lambda-3)(\lambda-5)$$

The **eigenvalues** are -1, 3 and 5 and therefore this form is **indefinite**.

Example 8.5.16. Let $\hat{q}$ be the **quadratic form** with **matrix** $A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{pmatrix}$.

Determine the type of $\hat{q}$.

The **characteristic polynomial** of A is $\chi_A(\lambda) =$

$$det \begin{pmatrix} 1-\lambda & -1 & -1 \\ -1 & 1-\lambda & 1 \\ -1 & 1 & 3-\lambda \end{pmatrix} =$$

$$(1-\lambda)(1-\lambda)(3-\lambda) + (-1)(1)(-1) + (-1)(-1)(1)$$

$$-(1-\lambda)(1)(1) - (-1)(-1)(3-\lambda) - (-1)(1-\lambda)(-1) =$$

$$-\lambda^3 + 5\lambda^2 - 7\lambda + 3 + 1 + 1 - 1 + \lambda - 3 + \lambda - 1 + \lambda =$$

$$-\lambda^3 + 5\lambda^2 - 4\lambda = -\lambda(\lambda - 1)(\lambda - 4)$$

The **eigenvalues** of this matrix are 0, 1 and 4 and so this form is **positive semidefinite**.

We now make use of what we have shown so far to classify **conic sections** and the **quadratic surfaces** in $\mathbb{R}^3$.

Classifying Conic Sections

We will now prove the following:

Theorem 8.5.4. *Let $\hat{p} = a_{11}X_1^2 + 2a_{12}X_1X_2 + a_{22}X_2^2 + b_1X_1 + b_2X_2 + c$ be a* **quadratic polynomial** *in two variables. Then $Z(\hat{p})$ is one of the following:*

i. the empty set;
ii. a single point;
iii. a line;
iv. a pair of parallel lines;
vi. a pair of intersecting lines;
vi. a parabola;
vii. an ellipse (circle); or
viii. a hyperbola.

Proof. Set $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$ and $B = (b_1 \; b_2)$. We may assume that $A \neq \mathbf{0}_{2\times 2}$ for otherwise $\hat{p}$ is linear and not a **quadratic polynomial**. There are two main cases as 1) $rank(A) = 1$; and 2) $rank(A) = 2$.

1) Assume that $rank(A) = 1$. By an **orthogonal change of variables** and a translation we can transform $\hat{p}(\boldsymbol{X})$ into $\hat{p}'(\boldsymbol{Y}) = \alpha_1 Y_1^2 - bY_2$ or $\alpha_1 Y_1^2 - c'$ where α_1 is the nonzero **eigenvalue** of A.

Assume that $\hat{p}' = \alpha_1 Y_1^2 - bY_2$ and that $\alpha_1 b > 0$. Then also $\frac{b}{\alpha_1} > 0$. Set $a = \sqrt{\frac{b}{\alpha_1}}$. Now

$$\hat{p}' = b[\frac{\alpha_1}{b}Y_1^2 - Y_2] = b[\frac{Y_1^2}{\frac{b}{\alpha_1}} - Y_1] = b[\frac{Y_1^2}{a^2} - Y_2]$$

In this case $Z(\hat{p}')$ is a **parabola** with vertical axis of symmetry and opens up in (Y_1, Y_2) space.

When $\alpha_1 b < 0$ set $a = \sqrt{-\frac{b}{\alpha_1}}$. Then

$$\hat{p}' = (-b)[-\frac{\alpha_1}{b}Y_1^2 + Y_2] = -b[\frac{Y_1^2}{a^2} + Y_2]$$

which is again a **parabola** with vertical axis of symmetry which opens down.

Now suppose that $\hat{p}' = \alpha_1 Y_1^2 - c'$. Notice in this case the quadratic form $\hat{p}$ is equivalent to a **quadratic polynomial** in a single variable. Now

$$\hat{p}' = \alpha_1[Y_1^2 - \frac{c'}{\alpha_1}]$$

if $\frac{c'}{\alpha_1} < 0$ then $Z(\hat{p}')$ and $Z(\hat{p})$ are empty. If $c' = 0$ then we obtain a single line, while if $\frac{c'}{\alpha_1} > 0$ then $Z(\hat{p}')$ consists of the two parallel lines $Y_1 = \sqrt{\frac{c'}{\alpha_1}}$ and $Y_1 = -\sqrt{\frac{c'}{\alpha_1}}$

2) Assume that $rank(A) = 2$. Let the **eigenvalues** of A be α_1 and α_2 with $\alpha_1 \geq \alpha_2$. By an **orthogonal change of variables** and a translation we can transform $\hat{p}$ into $\hat{p}'(\boldsymbol{Y}) = \alpha_1 Y_1^2 + \alpha_2 Y_2^2 - c'$ where α_1, α_2 are the **eigenvalues** of the matrix A. Several subcases occur.

i. α_1 and α_2 have the same sign ($\alpha_1\alpha_2 > 0$). In this situation the **quadratic part** $\hat{q}'$ of $\hat{p}'$ is either **positive definite** or **negative definite.** Suppose α_1 and α_2 are both positive and c' is negative. Then there are no solutions to $\alpha_1 Y_1^2 + \alpha_2 Y_2^2$ and $Z(\hat{q}')$, whence $Z(\hat{q})$ is empty.

When $c' = 0$ there is a unique point in $Z(\hat{p}')$ namely, $(Y_1, Y_2) = (0, 0)$.

Suppose $c' > 0$. Set $a = \sqrt{\frac{c'}{\alpha_1}}, b = \sqrt{\frac{c'}{\alpha_2}}$. Then

$$\hat{p}' = c'[\frac{\alpha_1}{c'}Y_1^2 + \frac{\alpha_2}{c'}Y_2^2 - 1] = c'[\frac{Y_1^2}{a^2} + \frac{Y_2^2}{b^2} - 1]$$

and in this case $Z(\hat{p}')$ is an **ellipse** (a circle when $\alpha_1 = \alpha_2$).

ii) When both α_1 and α_2 are negative the analysis is exactly the same.

iii) Assume that the **quadratic part** $\hat{q}'$ of $\hat{p}'$ is indefinite so that $\alpha_1 > 0$ and $\alpha_2 < 0$. There are three cases to consider: $c' > 0, c' = 0, c' < 0$.

Suppose $c' > 0$. Set $a = \sqrt{\frac{c'}{\alpha_1}}, b = \sqrt{-\frac{c'}{\alpha_2}}$. Then

$$\hat{p}' = c'[\frac{Y_1^2}{a^2} - \frac{Y_2^2}{b^2} - 1]$$

and $Z(\hat{p}')$ is a **hyperbola**.

When $c' = 0$ we obtain the two intersecting lines

$$Y_2 = \sqrt{\frac{\alpha_1}{\alpha_2}}Y_1, Y_2 = -\sqrt{\frac{\alpha_1}{\alpha_2}}Y_1$$

Finally, when $c' < 0$, set $a = \sqrt{-\frac{c'}{\alpha_1}}, b = \sqrt{\frac{c'}{\alpha_2}}$. Then

$$\hat{p}' = c'[\frac{Y_2^2}{b^2} - \frac{Y_1^2}{a^2} - 1]$$

and we again obtain a **hyperbola**. □

Remark 8.9. In the course of proving **Theorem** (8.5.4) we demonstrated that any **quadratic polynomial** $\hat{p}$ in two variables can be transformed by a **orthogonal change of variable** and a translation into one of the following forms:

1. aX_1^2.
2. $aX_1^2 + c$ with $ac < 0$.
3. $aX_1^2 + c$ with $ac > 0$.
4. $aX_1^2 + bX_2$.
5. $aX_1^2 + bX_2^2$ with $ab > 0$.
6. $aX_1^2 + bX_2^2$ with $ab < 0$.
7. $aX_1^2 + bX_2^2 + c$ with $ab, ac > 0$.
8. $aX_1^2 + bX_2^2 + c$ with $ab > 0, ac < 0$.
9. $aX_1^2 + bX_2^2 + c$ with $ab < 0, ac < 0$.

We refer to these as the ***standard forms*** of a **quadratic polynomial** in two variables.

Example 8.5.17. Determine which **conic section** is obtained from the **quadratic polynomial**

$$\hat{p} = 4X_1^2 + 4X_2^2 + 10X_1X_2 - 7\sqrt{2}X_1 - 11\sqrt{2}X_2 - 20$$

The **quadratic part** of $\hat{p}$ is $4X_1^2 + 4X_2^2 + 10X_1X_2$ which has **matrix** $A = \begin{pmatrix} 4 & 5 \\ 5 & 4 \end{pmatrix}$.

This matrix has **characteristic polynomial**

$$\chi_A(\lambda) = \lambda^2 - 8\lambda - 9 = (\lambda - 9)(\lambda + 1)$$

The **eigenvalues** are 9 and -1. The vectors

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

are **eigenvectors** . Notice they are **orthogonal**. By **normalizing** and taking the resulting vectors as columns we obtain the **orthogonal matrix** $Q = \begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}$

Set $\boldsymbol{Y} = \begin{pmatrix}Y_1\\Y_2\end{pmatrix} = Q^{Tr}\boldsymbol{X} = Q^{-1}\boldsymbol{X}$ so that $\boldsymbol{X} = Q\boldsymbol{Y}$. Then

$$\hat{p} = (Q\boldsymbol{Y})^{Tr}A(Q\boldsymbol{Y}) + (-7\sqrt{2} \ \ -11\sqrt{2})(Q\boldsymbol{Y}) - 20 =$$

$$\boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} + (-7\sqrt{2} \ \ -11\sqrt{2})\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}\boldsymbol{Y} - 20 =$$

$$(Y_1 \ Y_2)\begin{pmatrix}9 & 0\\0 & -1\end{pmatrix}\begin{pmatrix}Y_1\\Y_2\end{pmatrix} + (-18 \ \ -4)\begin{pmatrix}Y_1\\Y_2\end{pmatrix} - 20 =$$

$$9Y_1^2 - Y_2^2 - 18Y_1 - 4Y_2 - 20 =$$

$$9(Y_1^2 - 2Y_1 + 1) - (Y_2^2 + 4Y_2 + 4) - 9 + 4 - 20 =$$

$$9(Y_1 - 1)^2 - (Y_2 + 2)^2 - 25$$

Set $Z_1 = Y_1 - 1, Z_2 = Y_2 + 2, \boldsymbol{Z} = \begin{pmatrix}Z_1\\Z_2\end{pmatrix}$. Then

$$\hat{p} = 9Z_1^2 - Z_2^2 - 25 = \hat{p}'(\boldsymbol{Z}).$$

After factoring out 25 we get

$$\hat{p}'(\boldsymbol{Z}) = 25[\frac{Z_1^2}{(\frac{5}{3})^2} - \frac{Z_2^2}{5^2} - 1]$$

$Z(\hat{p}')$,whence $Z(\hat{p})$, is a **hyperbola**.

Classifying Quadratic Surfaces in $\mathbb{R}^3$

Suppose $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + C$ is a **quadratic polynomial** of three variables X_1, X_2 and X_3. If the **rank** of A is one we can show via a change of variables the polynomial can be transformed into a polynomial in one or two variables. Therefore in classifying the **quadratic surfaces** we may assume that the **symmetric matrix** A has **rank** at least two. We prove the following:

Theorem 8.5.5. *Let $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + C$ be a **quadratic polynomial** in three variables X_1, X_2, X_3. Assume the **symmetric matrix** A has **rank** at least two and that $\hat{p}$ cannot be transformed into a polynomial in two variables. Then $Z(\hat{p})$ is one of the following:*

i. empty;
ii. a single point;
iii. an ellipsoid (sphere);
iv. an elliptic cone;
v. a hyperboloid of one sheet;
vi. a hyperboloid of two sheets;
vii. an elliptic paraboloid; or
viii. a hyperbolic paraboloid.

Proof. By making use of an **orthogonal change of variable**, if necessary, we may assume that the **quadratic part** $\hat{q}$ of $\hat{p}$ is diagonal - there are no cross product terms.

There are two major case divisions: 1) $rank(A) = 2$ and 2) $rank(A) = 3$.

Case 1) $rank(A) = 2$. Let the nonzero **eigenvalues** of A be α_1, α_2 with the notation chosen so that if $\alpha_1\alpha_2 < 0$ then α_1 is positive. According to our assumption $\hat{p}$ has the form

$$\alpha_1 X_1^2 + \alpha_2 X_2^2 + b_1X_1 + b_2X_2 + b_3X_3 + C$$

Note that since we are assuming that $\hat{p}$ cannot be transformed into a polynomial in two variables we may assume that $b_3 \neq 0$. Set

$$Y_1 = X_1 + \frac{b_1}{2\alpha_1}, Y_2 = X_2 + \frac{b_2}{2\alpha_2}, C' = C - \frac{b_1^2}{4\alpha_1} - \frac{b_2^2}{4\alpha_2}$$

Also, if $b_3 < 0$ set $b = -b_3$ and $Y_3 = X_3 + \frac{C'}{b_3}$. On the other hand, if $b_3 > 0$ set $b = b_3$ and $Y_3 = -X_3 + \frac{C'}{b_3}$ Then

$$\alpha_1 X_1^2 + \alpha_2 X_2^2 + b_1X_1 + b_2X_2 + b_3X_3 + C = \alpha_1 Y_1^2 + \alpha_2 Y_2^2 - bY_3 = \hat{p}'(\boldsymbol{Y})$$

There are two subdivisions: (a) $\alpha_1\alpha_2 > 0$ and (b) $\alpha_1\alpha_2 < 0$.

Subcase (a). Notice that according to the definition of b that both $\alpha_1 b > 0$ and $\alpha_2 b > 0$. Set $d = \sqrt{\frac{b}{\alpha_1}}, e = \sqrt{\frac{b}{\alpha_2}}$. Then

$$\hat{p}' = b[\frac{Y_1^2}{d^2} + \frac{Y_2^2}{e^2} - Y_3]$$

and consequently $Z(\hat{p}')$ is an **elliptic paraboloid**.

Subcase (b) Assume that α_1 and α_2 have different signs (so by our assumption $\alpha_1 > 0$). It is still the case that $\alpha_1 b > 0$ but now $\alpha_2 b < 0$. Set $d = \sqrt{\frac{b}{\alpha_1}}$ and $e = \sqrt{-\frac{b}{\alpha_2}}$ Now

$$\hat{p}' = b[\frac{Y_1^2}{d^2} - \frac{Y_2^2}{e^2} - Y_3]$$

and $Z(\hat{p}')$ is a **hyperbolic paraboloid**.

Case 2) $rank(A) = 3$. Let $\alpha_1, \alpha_2, \alpha_3$ be the **eigenvalues** of A. Now we can apply **Theorem** (8.5.1) and we may assume that $\hat{p} = \alpha_1 X_1^2 + \alpha_2 X_2^2 + \alpha_3 X_3^2 + C$. There are two subcases to consider: (a) all the α_i have the same sign (A is **positive or negative definite**); and (b) there are both positive and negative **eigenvalues** (A is **indefinite**).

Subcase (a) If C has the same sign as the α_i then $Z(\hat{p})$ is empty. If $C = 0$ there is a single point. If $C\alpha_i < 0$ set $d = \sqrt{-\frac{C}{\alpha_1}}, e = \sqrt{-\frac{C}{\alpha_2}}, f = \sqrt{-\frac{C}{\alpha_3}}$. Then

$$\hat{p} = (-C)[\frac{X_1^2}{d^2} + \frac{X_2^2}{e^2} + \frac{X_3^2}{f^2} - 1]$$

and $Z(\hat{p})$ is an **ellipoid** (a sphere if $\alpha_1 = \alpha_2 = \alpha_3$).

Subcase (b) Since two of $\alpha_1, \alpha_2, \alpha_3$ have the same sign we can assume that the notation is chosen so that α_1 and α_2 have the same sign and α_3 has a different sign. There are three further divisions to consider: i. $C = 0$, ii. $\alpha_1 C > 0$ and ii. $\alpha_1 C < 0$.

i. In this case $Z(\hat{p})$ is an elliptic cone.

ii. In this case $\alpha_2 C$ is also positive and $\alpha_3 C < 0$. Set $a = \sqrt{\frac{C}{\alpha_1}}, b = \sqrt{\frac{C}{\alpha_2}}$ and $c = \sqrt{-\frac{C}{\alpha_3}}$. Then

$$\hat{p} = (-C)[\frac{X_3^2}{c^2} - \frac{X_1^2}{a^2} - \frac{X_2^2}{b^2} - 1].$$

$Z(\hat{p})$ is a **hyperboloid of two sheets**.

iii. In this case $\alpha_2 C$ is also negative and $\alpha_3 C > 0$. Set $a = \sqrt{-\frac{C}{\alpha_1}}, b = \sqrt{-\frac{C}{\alpha_2}}, c = \sqrt{\frac{C}{\alpha_3}}$. Then

$$\hat{p} = (-C)[\frac{X_1^2}{a^2} + \frac{X_2^2}{b^2} - \frac{X_3^2}{c^2} - 1]$$

Thus, $Z(\hat{p})$ is a **hyperboloid of one sheet**. □

Remark 8.10. In the course of proving **Theorem** (8.5.5) we have demonstrated that using a **orthogonal change of variable** and a translation we can transform any **quadratic polynomial** in three variables (which cannot be reduced to a polynomial in two variables) into one of the following forms:

1. $a_1X_1^2 + a_2X_2^2 + bX_3, a_1a_2 > 0$.
2. $a_1X_1^2 + a_2X_2^2 + bX_3, a_1a_2 < 0$.
3. $a_1X_1^2 + a_2X_2^2 + a_3X_3^2, a_1a_2, a_1a_3, a_2a_3 > 0$.
4. $a_1X_1^2 + a_2X_2^2 + a_3X_3^2, a_1a_2 > 0, a_1a_3 < 0, a_2a_3 < 0$.
5. $a_1X_1^2 + a_2X_2^2 + a_3X_3^2 + c, a_1a_2, a_1a_3, a_2a_3 > 0$.
6. $a_1X_1^2 + a_2X_2^2 + a_3X_3^2 + c, a_1a_2 > 0, a_1a_3 < 0, a_2a_3 < 0$.

We refer to these as the ***standard forms*** of a **quadratic polynomial** in three variables.

Example 8.5.18. Let

$$\hat{p} = 4X_1^2 + 5X_2^2 - 12X_1X_2 + 8X_1X_3 - 4X_2X_3 + 9X_1 - 18X_2 + 18X_3$$

Determine $Z(\hat{p})$.

Set $\hat{q} = 4X_1^2 + 5X_2^2 - 12X_1X_2 + 8X_1X_3 - 4X_2X_3$, the **quadratic part** of $\hat{p}$. Further, let A be the **matrix** of $\hat{q}$ so that $A = \begin{pmatrix} 4 & -6 & 4 \\ -6 & 5 & -2 \\ 4 & -2 & 0 \end{pmatrix}$. Finally, set $B = (9\ -18\ 18)$.

Then

$$\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X}$$

where $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$.

The **characteristic polynomial** of A is $-\lambda^3+9\lambda^2-36\lambda = -\lambda(\lambda+3)(\lambda-12)$. Thus, A has **eigenvalues** are 0, -3 and 12. The following vectors are **eigenvectors** (**normalized** to have **unit length**):

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

Let $Q = \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}$ so that $Q^{Tr}AQ = diag(0, -3, 12)$.

Set $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$ and $B' = BQ = (9\ \ -18\ 18)\begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix} = (3\ 12\ 24)$.

Substituting $\boldsymbol{X} = Q\boldsymbol{Y}$ we obtain

$$\hat{p}(\boldsymbol{X}) = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} + B(Q\boldsymbol{Y}) =$$

$$\boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} + B'\boldsymbol{Y} =$$

$$-3Y_2^2 + 12Y_3^2 + 3Y_1 + 12Y_2 + 24Y_3 =$$

$$3Y_1 - 3(Y_2^2 - 4Y_2) + 12(Y_3^2 + 2Y_2) =$$

$$3Y_1 - 3(Y_2^2 - 4Y_2 + 4) + 12(Y_3^2 + 2Y_3 + 1) - 12 + 12 =$$

$$3Y_1 - 3(Y_2 - 2)^2 + 12(Y_3 + 1)^2$$

Set $Z_1 = Y_1, Z_2 = Y_2 - 2, Z_3 = Y_3 + 1$ then

$$\hat{p} = 3Z_1 - 3(Z_2)^2 + 12(Z_3)^2 = 3[Z_1 - Z_2^2 + \frac{(Z_3)^2}{(\frac{1}{2})^2}] = \hat{p}'(\boldsymbol{Z})$$

$Z(\hat{p}')$, whence $Z(\hat{p})$, is a **hyperbolic paraboloid**.

Remark 8.11. Once we determined that the **eigenvalues** of the matrix A were 0, -3 and 12 we only needed to show that the polynomial was not reducible to two variables and we could conclude that it is was a **hyperbolic paraboloid**.

What You Can Now Do

1. Given a **quadratic form** $\hat{q}$ write down the **matrix** of $\hat{q}$.

2. Given a **quadratic form** $\hat{q}$ classify it as **positive definite, positive semidefinite, negative definite, negative semidefinite or indefinite**.

3. Given a **quadratic form** make an **orthogonal change of variable** to eliminate cross product terms.

4. Given a **quadratic polynomial** $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + C$ in two variables use an **Orthogonal change of variable** and a translation to put the polynomial into **standard form** and identify the **conic section** $Z(\hat{p})$.

5. Given a **quadratic polynomial** $\hat{p} = \boldsymbol{X}^{Tr} A \boldsymbol{X} + B\boldsymbol{X} + C$ in three variables use an **orthogonal change of variable** and a translation to put the polynomial into **standard form** and identify the **quadratic surface** $Z(\hat{p})$.

Method (How To Do It)

Method 8.5.1. Given a **quadratic form** $\hat{q}$ write down the **matrix** of $\hat{q}$.

Set a_{ii} equal to the coefficient of X_i^2 in $\hat{q}$ and $a_{ij} = a_{ji}$ equal to one half the coefficient of $X_i X_j$ in $\hat{q}$ for $i < j$.

Example 8.5.19. Write down the **matrix** of the **quadratic form**

$$\hat{q} = -3X_1^2 + 2X_2^2 - 5X_3^2 + X_1X_2 - 6X_1X_3 + 7X_2X_3$$

The matrix is $A = \begin{pmatrix} -3 & \frac{1}{2} & -3 \\ \frac{1}{2} & 2 & \frac{7}{2} \\ -3 & \frac{7}{2} & -5 \end{pmatrix}$

Method 8.5.2. Given a **quadratic form** $\hat{q}$ classify it as **positive definite, positive semidefinite, negative definite, negative semidefinite or indefinite**.

Determine the **eigenvalues** of the matrix A of $\hat{q}$ using **Method** (7.1.2). If all the **eigenvalues** are positive then $\hat{q}$ is **positive definite**. If all the **eigenvalues** are non-negative and some **eigenvalue** is zero then the form is **positive semidefinite**. If all the **eigenvalues** are negative then the form is **negative definite**. If all the **eigenvalues** are non-positive and some **eigenvalue** is zero the form is **negative semidefinite**. Finally, if there are both positive and negative **eigenvalues** then the form is **indefinite**.

Example 8.5.20. Classify the quadratic forms given in (a) - (c) below.

(a) $\hat{q} = -3X_1^2 + 5X_2^2 + 6X_1X_2$. The matrix of $\hat{q}$ is

$$A = \begin{pmatrix} -3 & 3 \\ 3 & 5 \end{pmatrix}$$

The **characteristic polynomial** of A is

$$\chi_A(\lambda) = \lambda^2 - 2\lambda - 24 = (\lambda - 6)(\lambda + 4)$$

Therefore the **eigenvalues** are 6 and -4 and $\hat{q}$ is **indefinite**.

(b) $\hat{q} = 3X_1^2 + 3X_2^2 + 5X_3^2 - 4X_1X_3 + 4X_2X_3$. The matrix of $\hat{q}$ is

$$A = \begin{pmatrix} 3 & 0 & -2 \\ 0 & 3 & 2 \\ -2 & 2 & 5 \end{pmatrix}$$

The **characteristic polynomial** of A is

$$\chi_A(\lambda) = (3-\lambda)(3-\lambda)(5-\lambda) - (3-\lambda)(2)(2) - (-2)(3-\lambda)(-2) =$$

$$-\lambda^3 + 11\lambda^2 - 39\lambda + 45 - 12 + 4\lambda - 12 + 4\lambda =$$

$$-\lambda^3 + 11\lambda - 31\lambda + 21 =$$

$$-(\lambda - 1)(\lambda - 3)(\lambda - 7)$$

The **eigenvalues** are 1, 3 and 7 and the form is **positive definite**.

(c) $\hat{q} = -2X_1^2 - 2X_2^2 - 6X_3^2 + 4X_1X_2 + 4X_1X_3 - 4X_2X_3$. The matrix of $\hat{q}$ is

$$A = \begin{pmatrix} -2 & 2 & 2 \\ 2 & -2 & -2 \\ 2 & -2 & -6 \end{pmatrix}$$

The **characteristic polynomial** of A is

$$\chi_A(\lambda) = (-2-\lambda)(-2-\lambda)(-6-\lambda) + (2)(-2)(2) + (2)(2)(-2)$$

$$-(-2-\lambda)(-2)(-2) - (2)(2)(-6-\lambda) - (2)(-2-\lambda)(2) =$$

$$-\lambda^3 - 10\lambda^2 - 28\lambda - 24 - 8 - 8 + 8 + 4\lambda + 24 + 4\lambda + 8 + 4\lambda =$$

$$-\lambda^3 - 10\lambda^2 - 16\lambda = -\lambda(\lambda^2 + 10\lambda + 16) =$$

$$-\lambda(\lambda + 2)(\lambda + 8)$$

Thus, the **eigenvalues** are 0, -2 and -8 and this form is **negative semidefinite**.

Method 8.5.3. Given a **quadratic form** make an **orthogonal change of variable** to eliminate cross product terms.

Let A be the **matrix** of $\hat{q}$. Find an **orthogonal matrix** Q which diagonalizes A by **Method** (8.4.2). Set $D = diag(\alpha_1, \alpha_2, \ldots, \alpha_n) = Q^{Tr}AQ$ and $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$. Then

$$\hat{q}(\boldsymbol{X}) = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} = \boldsymbol{Y}^{Tr}D\boldsymbol{Y}$$

has no cross product terms.

Example 8.5.21. Find an **orthogonal change of variables** which eliminates the cross products terms from the **quadratic forms** of **Example** (8.5.20) (a) and (b).

(a) $\hat{q} = -3X_1^2 + 5X_2^2 + 6X_1X_2$. We saw that the **eigenvalues** are -4 and 6. Corresponding are: $\begin{pmatrix} \frac{3}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} \end{pmatrix}$ and $\begin{pmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{pmatrix}$. Set $Q = \begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix}$. If $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$ then

$$\hat{q} = -4Y_1^2 + 6Y_2^2$$

(b) $\hat{q} = 3X_1^2 + 3X_2^2 + 5X_3^2 - 4X_1X_3 + 4X_2X_3$. The **eigenvalues** of the matrix of this **quadratic form** are 1, 3 and 7. Corresponding **eigenvectors** are:

$$\boldsymbol{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$$

Set $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3)$, an **orthogonal matrix** and $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$. Then

$$\hat{q} = Y_1^2 + 3Y_2^2 + 7Y_3^2$$

Method 8.5.4. Given a **quadratic polynomial** $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + C$ in two variables use an **orthogonal change of variable** and a translation to put the polynomial into **standard form** and identify the **conic section** $Z(\hat{p})$.

Let the **eigenvalues** of A be α_1 and α_2. First find an **orthogonal matrix** Q such that $Q^{Tr}AQ = D = diag(\alpha_1, \alpha_2)$ is diagonal as in **Method** (8.5.3). Set $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$ and $B' = BQ$. Then

$$\hat{p} = \boldsymbol{Y}^{Tr} D\boldsymbol{Y} + B'\boldsymbol{Y} + C = \alpha_1 Y_1^2 + \alpha_2 Y_2^2 + b_1' Y_1 + b_2' Y_2 + C$$

If α_1 and α_2 are both non-zero set $Z_1 = Y_1 + \frac{b_1'}{2\alpha_1}, Z_2 = Y_2 + \frac{b_2'}{2\alpha_2}$ and $C' = C - \frac{(b_1')^2}{4\alpha_1} - \frac{(b_2')^2}{4\alpha_2}$. Then

$$\hat{p} = \alpha_1 Z_1^2 + \alpha_2 Z_2^2 + C'$$

Suppose $\alpha_2 = 0$. Set $Z_1 = Y_1 + \frac{b_1'}{2\alpha_1}, Z_2 = Y_2 + \frac{C - \frac{(b_1')^2}{4\alpha_1}}{b_2'}$. Then

$$\hat{p} = \alpha_1 Z_1^2 + b_2' Z_2'$$

Example 8.5.22. Let

$$\hat{p} = -3X_1^2 - 4X_1X_2 + \frac{52}{\sqrt{5}}X_1 + \frac{15}{\sqrt{5}}X_2 - 36$$

Find the **standard form** of $\hat{p}$ and identify the **conic section** $Z(\hat{p})$.

Let $\hat{q} = -3X_1^2 - 4X_1X_2$, the **quadratic part** of $\hat{p}$. Let A be the **matrix** of $\hat{q}$ so that $A = \begin{pmatrix} -3 & -2 \\ -2 & 0 \end{pmatrix}$. Also, set $B = (\frac{52}{\sqrt{5}}\ \frac{16}{\sqrt{5}})$ so that

$$\hat{p} = \boldsymbol{X}^{Tr} A\boldsymbol{X} + B\boldsymbol{X} - 36$$

The **characteristic polynomial** of A is

$$\chi_A(\lambda) = \lambda^2 + 3\lambda - 4 = (\lambda - 1)(\lambda + 4)$$

and therefore A has **eigenvalues** are -4 and 1. The following are **eigenvectors** for the **eigenvalues** , respectively:

$$\boldsymbol{v}_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \boldsymbol{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}$$

Set $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2)$ so that Q is an **orthogonal matrix** and $Q^{Tr}AQ = diag(-4, 1)$. Set $B' = BQ =$

$$(\frac{52}{\sqrt{5}}\ \frac{16}{\sqrt{5}}) \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix} = (24\ 4)$$

Set $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = Q^{Tr}\boldsymbol{X}$ so that $\boldsymbol{X} = Q\boldsymbol{Y}$. Then

$$\hat{p} = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} + B(Q\boldsymbol{Y}) - 36 = -4Y_1^2 + Y_2^2 + 24Y_1 + 4Y_2 - 36$$

Set $Z_1 = Y_1 - 3, Z_2 = Y_2 + 2$. Then

$$\hat{p} = -4Z_1^2 + Z_2^2 + 36 - 4 - 36 = Z_1^2 - 4Z_2^2 - 4 = 4[-Z_1^2 + \frac{Z_2^2}{2^2} - 1]$$

The **standard form** of $\hat{p}$ is $4[-Z_1^2 + \frac{Z_2^2}{2^2} - 1]$ and $Z(\hat{p})$ is a **hyperbola**.

Example 8.5.23. Find the **standard form** of the **quadratic polynomial** $\hat{p} = 8X_1^2 + 16X_1X_2 + 8X_2^2 - 14\sqrt{2}X_1 - 18\sqrt{2}X_2$ and identify the **conic section** $Z(\hat{p})$.

Let $\hat{q} = 8X_1^2 + 16X_1X_2 + 8X_2^2$, the **quadratic part** of $\hat{p}$. Let A be the **matrix** of $\hat{q}$ so that $A = \begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix}$ and set $B = (-14\sqrt{2} \;\; -18\sqrt{2})$ so that $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X}$. The **characteristic polynomial** of A is

$$\chi_A(\lambda) = \lambda^2 - 16\lambda = \lambda(\lambda - 16)$$

and therefore the **eigenvalues** are 16 and 0. The following are **eigenvectors** of **unit length** for the the respective **eigenvalues**:

$$\boldsymbol{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Set $Q = (\boldsymbol{v}_1 \; \boldsymbol{v}_2)$ so that Q is an **orthogonal matrix** and $Q^{Tr}AQ = diag(16, 0)$. Set $B' = BQ =$

$$(-14\sqrt{2} \;\; -18\sqrt{2}) \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = (-32 \;\; -4)$$

Set $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = Q^{Tr}\boldsymbol{X}$ so that $\boldsymbol{X} = Q\boldsymbol{Y}$. Then

$$\hat{p} = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} + B(Q\boldsymbol{Y}) = 16Y_1^2 - 32Y_1 - 4Y_2$$

Set $Z_1 = Y_1 - 1, Z_2 = Y_2 + 4$. Then

$$\hat{p} = 16Z_1^2 - 4Z_2 = 4[4Z_1^2 - Z_2]$$

The **standard form** of $\hat{p}$ is $4[4Z_1^2 - Z_2]$ and $Z(\hat{p})$ is a **parabola**.

Method 8.5.5. Given a **quadratic polynomial** $\hat{p} = \boldsymbol{X}^{Tr} A\boldsymbol{X} + B\boldsymbol{X} + C$ in three variables use an **orthogonal change of variable** and a translation to put the polynomial into **standard form** and identify the **quadratic surface** $Z(\hat{p})$.

If A has **rank** one then $\hat{p}$ is reducible to a **quadratic polynomial** in one or two variables so we assume that $rank(A) \geq 2$. Then at least two of the **eigenvalues** of A are non-zero. Use **Method** (7.1.2) to find the **eigenvalues** $\alpha_1, \alpha_2, \alpha_3$ with the notation chosen so that α_1 and α_2 are non-zero and, if all three are non-zero so that α_1 and α_2 have the same sign.

First find an **orthogonal matrix** Q such that $Q^{Tr} AQ = D = diag(\alpha_1, \alpha_2, \alpha_3)$ is diagonal as in **Method** (8.4.3). Set $\boldsymbol{Y} = Q^{Tr}\boldsymbol{X}$ and $B' = BQ$. Then

$$\hat{p} = \boldsymbol{Y}^{Tr} D\boldsymbol{Y} + B'\boldsymbol{Y} + C = \alpha_1 Y_1^2 + \alpha_2 Y_2^2 + \alpha_3 Y_3^2 + b_1' Y_1 + b_2' Y_2 + b_3' Y_3 + C$$

Suppose first that $\alpha_1, \alpha_2, \alpha_3$ are all nonzero. Then set $Z_1 = Y_1 + \frac{b_1'}{2\alpha_1}, Z_2 = Y_2 + \frac{b_2'}{2\alpha_2}, Z_3 = Y_3 + \frac{b_3'}{2\alpha_3}$ and $C' = C - \frac{(b_1')^2}{4\alpha_1} - \frac{(b_2')^2}{4\alpha_2} - \frac{(b_3')^2}{4\alpha_3}$. Then

$$\hat{p} = \alpha_1 Z_1^2 + \alpha_2 Z_2^2 + \alpha_3 Z_3^2 + C'$$

is the **standard form** of $\hat{p}$.

Suppose all the α_i have the same sign. If C' has the same sign as well then there are no solutions. If $C' = 0$ then there is a unique point in $Z(\hat{p})$. Finally, if C' has the opposite sign to $\alpha_1, \alpha_2, \alpha_3$ then $Z(\hat{p})$ is a **ellipsoid** (a sphere if $\alpha_1 = \alpha_2 = \alpha_3$).

Suppose α_3 has a different sign from α_1 and α_2. If $\alpha_1 C'$ is positive then $Z(\hat{p})$ is a **hyperboloid of two sheets**, whereas if $\alpha_1 C'$ is negative then $Z(\hat{p})$ is a **hyperboloid of one sheet**. If $C' = 0$ then $Z(\hat{p})$ is a —tr**elliptic cone**.

Assume now that $\alpha_3 = 0$. Now set $Z_1 = Y_1 + \frac{b_1'}{2\alpha_1}, Z_2 = Y_2 + \frac{b_2'}{2\alpha_2}, Z_3 = Y_3 - \frac{1}{b_3'}[\frac{(b_1')^2}{4\alpha_1} + \frac{(b_2')^2}{4\alpha_2} - C]$. Then

$$\hat{p} = \alpha_1 Z_1^2 + \alpha_2 Z_2^2 + b_3' Z_3$$

is the **standard form** of $\hat{p}$.

If α_1 and α_2 have the same sign then $Z(\hat{p})$ is an elliptic paraboloid and if α_1, α_2 have different signs then $Z(\hat{p})$ is a hyperbolic paraboloid.

Example 8.5.24. Let $\hat{p}$ be the **quadratic polynomial** $7X_1^2+6X_2^2+5X_3^2-4X_1X_2-4X_2X_3+30X_1-12X_2+24X_3+62$. Find the **standard form** of $\hat{p}$ and identify the **quadratic surface** $Z(\hat{p})$.

Set $\hat{q} = X_1^2+6X_2^2+5X_3^2-4X_1X_2-4X_2X_3$ and let A be the **matrix** of $\hat{q}$ so that $A = \begin{pmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{pmatrix}$. Also, set $B = \begin{pmatrix} 30 & -12 & 24 \end{pmatrix}$ so that $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X} + 62$. The **characteristic polynomial** of A is

$$\chi_A(\lambda) = (7-\lambda)(6-\lambda)(5-\lambda) - (7-\lambda)(-2)(-2) - (-2)(-2)(5-\lambda) =$$

$$-\lambda^3 + 18\lambda^2 - 99\lambda + 162 =$$

$$-(\lambda-3)(\lambda-6)(\lambda-9)$$

Thus, A has **eigenvalues** are 3, 6 and 9. The following are **eigenvectors** of **unit length** for the respective **eigenvalues**:

$$\boldsymbol{v}_1 = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$$

Set $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3)$ so that Q is an **orthogonal matrix** and $Q^{Tr}AQ = diag(3,6,9)$. Set $B' = BQ =$

$$\begin{pmatrix} 30 & -12 & 24 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} 18 & 0 & -36 \end{pmatrix}.$$

Set $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = Q^{Tr}\boldsymbol{X}$ so that $\boldsymbol{X} = Q\boldsymbol{Y}$. Then

$$\hat{p} = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} + B(Q\boldsymbol{Y}) = 3Y_1^2 + 6Y_2^2 + 9Y_3^2 + 18Y_1 - 36Y_3 + 62$$

Set $Z_1 = Y_1 + 3, Z_2 = Y_2$ and $Z_3 = Y_3 - 2$ Then

$$\hat{p} = 3Z_1^2 + 6Z_2^2 + 9Z_3^2 - 27 - 36 + 62 = 3Z_1^2 + 6Z_2^2 + 9Z_3^2 - 1$$

The **standard form** of $\hat{p}$ is $3Z_1^2 + 6Z_2^2 + 9Z_3^2 - 1$ and $Z(\hat{p})$ is an **ellipsoid**.

Example 8.5.25. Let $\hat{p} = 3X_1^2+3X_2^2+4X_3^2-6X_1X_2+4X_1X_3-4X_2X_3-\sqrt{2}X_1-\sqrt{2}X_2$. Find the **standard form** of $\hat{p}$ and identify the **quadratic surface** $Z(\hat{p})$.

Set $\hat{q} = 3X_1^2 + 3X_2^2 + 4X_3^2 - 6X_1X_2 + 4X_1X_3 - 4X_2X_3$, the **quadratic part** of $\hat{p}$, and let A be the **matrix** of $\hat{q}$ so that $A = \begin{pmatrix} 3 & -3 & 2 \\ -3 & 3 & -2 \\ 2 & -2 & 4 \end{pmatrix}$. Finally, set $B = \begin{pmatrix} -\sqrt{2} & -\sqrt{2} & 0 \end{pmatrix}$ so that $\hat{p} = \boldsymbol{X}^{Tr}A\boldsymbol{X} + B\boldsymbol{X}$. The **characteristic polynomial** of A is

$$\begin{gathered} \chi_A(\lambda) = (3-\lambda)(3-\lambda)(4-\lambda) + (-3)(-2)(2) + (2)(-3)(-2) \\ -(3-\lambda)(-2)(-2) - (-3)(-3)(4-\lambda) - (2)(3-\lambda)(2) = \\ -\lambda^3 + 10\lambda^2 - 16\lambda = \\ -(\lambda-8)(\lambda-2)\lambda \end{gathered}$$

Thus, the **eigenvalues** of A are 8, 2 and 0. The following are **eigenvectors** of **unit length** for the respective **eigenvalues**:

$$\boldsymbol{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

Set $Q = (\boldsymbol{v}_1\ \boldsymbol{v}_2\ \boldsymbol{v}_3)$ so that Q is an **orthogonal matrix** and $Q^{Tr}AQ = diag(8,2,0)$. Set $B' = BQ =$

$$\begin{pmatrix} -\sqrt{2} & -\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -2 \end{pmatrix}.$$

Set $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = Q^{Tr}\boldsymbol{X}$ so that $\boldsymbol{X} = Q\boldsymbol{Y}$. Then

$$\hat{p} = \boldsymbol{Y}^{Tr}(Q^{Tr}AQ)\boldsymbol{Y} + B(Q\boldsymbol{Y}) = 8Y_1^2 + 2Y_2^2 - 2Y_3 = 2[4Y_1^2 + Y_2^2 - Y_3]$$

The **standard form** of $\hat{p}$ is $2[4Y_1^2 + Y_2^2 - Y_3]$ and $Z(\hat{p})$ is an elliptic paraboloid.

Exercises

In exercises 1 - 4 write down the **matrix** of the given **quadratic form** $\hat{q}$.

1. $\hat{q} = 3X_1^2 + 4X_2^2 - 8X_1X_2$

2. $\hat{q} = -5X_1^2 + 3X_2^2 - X_1X_2$

3. $\hat{q} = -X_2^2 - 4X_3^2 + 2X_1X_2 - 6X_1X_3$

4. $\hat{q} = -X_1^2 + 3X_2^2 - 8X_3^2 + X_1X_2 - 4X_1X_3 + 5X_2X_3$

In exercises 5 - 12 classify the given **quadratic form** as **positive definite, positive semidefinite, negative definite, negative semidefinite or indefnite**.

5. $\hat{q} = 3X_1^2 + 3X_2^2 - 4X_1X_2$

6. $\hat{q} = 2X_1^2 - X_2^2 - 4X_1X_2$

7. $\hat{q} = -4X_1X_2 + 3X_2^2$

8. $\hat{q} = 3X_1^2 + 2X_2^2 + 4X_3^2 + 4X_1X_2 + 4X_1X_3$

9. $\hat{q} = 5X_1^2 + 5X_2^2 + 5X_3^2 - 2X_1X_2 - 2X_1X_3 - 2X_2X_3$

10. $\hat{q} = -2X_1^2 - 2X_2^2 - 5X_3^2 + 4X_1X_2 + 2X_1X_3 - 2X_2X_3$

11. $\hat{q} = -3X_1^2 - 6X_2^2 - 3X_3^2 + 4X_2X_3$

12. $\hat{q} = -X_1^2 - 5X_2^2 - 10X_3^2 + 4X_1X_2 + 6X_1X_3 - 12X_2X_3$

In exercises 13 - 16 find the **standard form** of the given **quadratic polynomial** $\hat{p}$ in the variables X_1 and X_2 and identify the **conic section** $Z(\hat{p})$.

13. $\hat{p} = 4X_1^2 + 36X_2^2 + 24X_1X_2 + 27X_1 - 9X_2$

14. $\hat{p} = -3X_1^2 - 4X_1X_2 - 4$

15. $\hat{p} = 37X_1^2 + 13X_2^2 - 18X_1X_2 - 16X_1 + 32X_2 - 20$

16. $\hat{p} = X_1^2 + 7X_2^2 + 8X_1X_2 + 30X_1 + 12X_2 + 27$

In exercises 17 - 24 find the **standard form** of the given **quadratic polynomial** $\hat{p}$ in the variables X_1, X_2, X_3 and identify the **quadratic surface** $Z(\hat{p})$.

17. $\hat{p} = 10X_1^2 + 10X_2^2 + 2X_3^2 - 18X_1X_2 - 72$

18. $\hat{p} = X_1^2 + X_2^2 + 2X_3^2 + 4X_1X_2 - 2X_1X_3 - 2X_2X_3 - 4$

19. $\hat{p} = 5X_1^2 + 5X_2^2 + 3X_3^2 + 2X_1X_2 + 6X_1X_3 + 6X_2X_3 + 4X_1 + 4X_2 + 10X_3 + \frac{4}{3}$

20. $\hat{p} = 15X_1^2 + 12X_2^2 + 36X_1X_2 + 12X_1X_3 + 24X_2X_3 - 8X_1 + 4X_2 + 8X_3$

21. $\hat{p} = X_1^2 + X_2^2 + X_3^2 - 4X_1 + 6X_2 + 2X_3 - 11$

22. $\hat{p} = 9X_1^2+9X_2^2+18X_3^3+6X_1X_2-12X_1X_3-12X_2X_3-24X_1-24X_2+24X_3+23$

23. $\hat{p} = -X_1^2 + 2X_2^2 + 7X_3^2 + 10X_1X_3 + 16X_2X_3 + 2X_1 + 2X_2 - 2X_3 - 8$

24. $\hat{p} = 2X_3^2 + 6X_1X_2 - 2X_1X_3 - 2X_2X_3 + 6X_1 + 6X_2 - 12X_3 + 27$

In exercises 25 - 30 answer true or false and give an explanation.

25. If $\hat{q}$ is a **quadratic form** then $\hat{q}(c\boldsymbol{v}) = c\hat{q}(\boldsymbol{v})$ for a scalar c.

26. If A is an $n \times n$ matrix then the function $\hat{q}$ defined by $\hat{q}(\boldsymbol{X}) = \boldsymbol{X}^T A\boldsymbol{X}$ is a **quadratic form**.

27. The **matrix** of a **quadratic form** is a **symmetric matrix**.

28. If a **quadratic form** $\hat{q}$ is **positive definite** then for every non-zero vector $\boldsymbol{x}$, $\hat{q}(\boldsymbol{x}) > 0$.

29. If A is a **symmetric matrix** then the **quadratic form** $\hat{q}$ given by $\hat{q}(\boldsymbol{x}) = \boldsymbol{X}^T A\boldsymbol{X}$ is **positive definite** if and only if the **eigenvalues** of A are all non-negative.

30. If A is an **invertible** 2×2 matrix and $c \neq 0$ is a scalar then the set $\{\boldsymbol{x} : \boldsymbol{x}^T A\boldsymbol{x} = c\}$ is either a circle or an ellipse.

Challenge Exercises (Problems)

1. Assume that $\hat{p} = \boldsymbol{X}^{Tr} A\boldsymbol{X} + B\boldsymbol{X} + C$ is a **quadratic polynomial** in three variables and that $rank(A) = 1$. Prove there is an **orthogonal change of variable** which reduces $\hat{p}$ to a quadratic polynomial in one or two variables.

2. Let $\hat{q} : \mathbb{R}^n \to \mathbb{R}$ be a function. Define $\langle\ ,\ \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \hat{q}(\boldsymbol{x} + \boldsymbol{y}) - \hat{q}(\boldsymbol{x}) - \hat{q}(\boldsymbol{y}) \tag{8.49}$$

Prove that if $\hat{q}$ is a **quadratic form** then the following hold:

a) For every pair of vectors $\boldsymbol{x}, \boldsymbol{y}$, $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{y}, \boldsymbol{x}\rangle$

b) For all vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}$, $\langle \boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}\rangle = \langle \boldsymbol{x}_1, \boldsymbol{y}\rangle + \langle \boldsymbol{x}_2, \boldsymbol{y}\rangle$

c) For all vectors $\boldsymbol{x}, \boldsymbol{y}$ and scalar c, $\langle c\boldsymbol{x}, \boldsymbol{y}\rangle = c\langle \boldsymbol{x}, \boldsymbol{y}\rangle$

3. Let $\hat{q} : \mathbb{R}^n \to \mathbb{R}$ be a function and assume that the following are satisfied:

i. $\hat{q}(c\boldsymbol{x}) = c^2\hat{q}(\boldsymbol{x})$ for every vector $\boldsymbol{x}$ and scalar c;
ii. $\langle \boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}\rangle = \langle \boldsymbol{x}_1, \boldsymbol{y}\rangle + \langle \boldsymbol{x}_2, \boldsymbol{y}\rangle$ for all vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}$; and
iii. $\langle c\boldsymbol{x}, \boldsymbol{y}\rangle = c\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all vectors $\boldsymbol{x}, \boldsymbol{y}$ and scalars c.

Set $a_{ij} = \langle \boldsymbol{e}_i, \boldsymbol{e}_j\rangle$ where $\boldsymbol{e}_i$ is the i^{th} **standard basis vector** and set $A =$

$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$. Prove that A is **symmetric** and for all vectors $\boldsymbol{x} \in \mathbb{R}^n$ that $\hat{q}(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{Tr}A\boldsymbol{x}$.

Quiz Solutions

1. $\begin{pmatrix} -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix}$. Not right, see **Method** (8.4.2).

2. $\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$. Not right, see **Method** (8.4.2).

3. $\begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$. Not right, see **Method** (8.4.2).

4. $\begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$. Not right, see **Method** (8.4.2).

8.6. How to Use It: Least Squares Approximation

In this section we will introduce the general least squares problem and demonstrate how to find a solution. We will then apply this to the problem of approximating data, in the form of a collection of ordered pairs, by a polynomial or exponential function of best fit.

Am I Ready for This Material

The following are essential to an understanding of the material introduced in this section:

linear system

inconsistent linear system

null space of a matrix

invertible matrix

transpose of a matrix

column space of a matrix

rank of a matrix

upper triangular matrix

QR factorization of a matrix

linear dependent sequence of vectors

dot product

orthogonal or perpendicular vectors

orthogonal complement of a **subspace** W of $\mathbb{R}^n$.

norm (length, magnitude) of a vector

orthonormal sequence of vectors

orthonormal basis of a subspace of $\mathbb{R}^n$

orthonormal matrix

The General Least Squares Problem

By the "general least squares problem" we mean the following:

Given an $m \times n$ matrix A and an $m-$vector $\boldsymbol{b}$ which is not in the **column space** of A, find a vector $\boldsymbol{x}$ such that $\parallel \boldsymbol{b} - A\boldsymbol{x} \parallel$, equivalently, $\parallel \boldsymbol{b} - A\boldsymbol{x} \parallel^2$, is as small as possible. This is referred to as a ***least squares problem*** since $\parallel \boldsymbol{b} - A\boldsymbol{x} \parallel^2$ is a sum of squares. A vector $\boldsymbol{x}$ that satisfies this is said to be a ***least squares solution***.

For any vector $\boldsymbol{x} \in \mathbb{R}^n$, the vector $A\boldsymbol{x}$ is in the **column space** of A. The first step in the solution to this problem is to identify the vector $A\boldsymbol{x}$. Immediately relevant to this is the **Best Approximation Theorem** which we state here:

Theorem (8.3.7)

Let W be a subspace of $\mathbb{R}^n$ and $\boldsymbol{u}$ an $n-$vector. Then for any vector $\boldsymbol{w} \in W, \boldsymbol{w} \neq Proj_W(\boldsymbol{u})$, $\parallel \boldsymbol{u} - Proj_W(\boldsymbol{u}) \parallel < \parallel \boldsymbol{u} - \boldsymbol{w} \parallel$.

Finding General Least Squares Solutions

Given an $m \times n$ matrix A and $\boldsymbol{b} \in \mathbb{R}^m$, set $W = col(A)$ and $\boldsymbol{b}' = Proj_W(\boldsymbol{b})$ and assume that $\boldsymbol{x}$ is a vector such that $A\boldsymbol{x} = \boldsymbol{b}'$.

Recall that the vector $\boldsymbol{b} - \boldsymbol{b}' = \boldsymbol{b} - A\boldsymbol{x}$ is in the **orthogonal complement**, $W^\perp$, to $W = col(A)$. This means that the vector $\boldsymbol{b} - \boldsymbol{b}' = \boldsymbol{b} - A\boldsymbol{x}$ is **orthogonal** to every column of the matrix A and therefore is in the **null space** of the **transpose** of A, A^{Tr}. This means that

$$A^{Tr}(\boldsymbol{b} - \boldsymbol{b}') = A^{Tr}(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{0}_n \tag{8.50}$$

An immediate consequence of (8.50) is that a vector $\boldsymbol{x}$ for which $A\boldsymbol{x} = \boldsymbol{b}' = Proj_W(\boldsymbol{b})$ satisfies the equation

$$A^{Tr}A\boldsymbol{x} = A^{Tr}\boldsymbol{b} \tag{8.51}$$

The equations which make up the **linear system** equivalent to the matrix equation (8.51) are referred to as the ***normal equations*** of $A\boldsymbol{x} = \boldsymbol{b}$.

We have shown that every **least squares solution** satisfies the **normal equations**. The converse is also true:

Theorem 8.6.1. *Assume that $\boldsymbol{x}$ satisfies the $A^{Tr}A\boldsymbol{x} = A^{Tr}\boldsymbol{b}$. Then $\boldsymbol{x}$ is a* **least squares solution** *to $A\boldsymbol{x} = \boldsymbol{b}$.*

Proof. Assume that $\boldsymbol{x}$ satisfies $A^{Tr}A\boldsymbol{x} = A^{Tr}\boldsymbol{b}$. Then

$$A^{Tr}A\boldsymbol{x} - A^{Tr}\boldsymbol{b} = A^{Tr}(A\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}_n \tag{8.52}$$

A consequence of (8.52) is that the vector $A\boldsymbol{x} - \boldsymbol{b}$ is in the **null space** of A^{Tr} and therefore **orthogonal** to every row of A^{Tr}, equivalently, every column of A. Since $A\boldsymbol{x} - \boldsymbol{b}$ is **orthogonal** to every column of A it follows that $A\boldsymbol{x} - \boldsymbol{b}$ is in the **orthogonal complement** of the **column space** of A, $col(A)$.

On the other hand, $A\boldsymbol{x}$ is in the **column space** of A and $\boldsymbol{b} = A\boldsymbol{x} + (\boldsymbol{b} - A\boldsymbol{x})$, the sum of a vector in $col(A)$ and a vector in $col(A)^{\perp}$. By **Theorem** (8.3.1) there are unique vectors $\boldsymbol{w} \in col(A)$ and $\boldsymbol{z} \in col(A)^{\perp}$ such that $\boldsymbol{b} = \boldsymbol{w} + \boldsymbol{z}$. Moreover, the vector $\boldsymbol{w} = Proj_{col(A)}(\boldsymbol{b})$. Thus, $A\boldsymbol{x} = Proj_{col(A)}(\boldsymbol{b})$ and is therefore a **least squares solution** to $A\boldsymbol{x} = \boldsymbol{b}$. □

Before proceeding to some examples we determine when a unique solution exists. Of course, this occurs precisely when the matrix $A^{Tr}A$ is **invertible**.

Theorem 8.6.2. *Let A be an $m \times n$ matrix and $\boldsymbol{b} \in \mathbb{R}^m$ a vector such that the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ has no solutions. Then the* **least squares problem** *for the system $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution if and only if the sequence of columns of the matrix A is* **linearly independent**. *In this case the unique solution is given by $\boldsymbol{x} = (A^{Tr}A)^{-1}A^{Tr}\boldsymbol{b}$.*

Proof. First assume that there is a unique solution. Then $A^{Tr}A$ is **invertible** which implies that the **rank** of $A^{Tr}A$ is n. In general if B and C are matrices which can be multiplied, then $rank(BC) \leq min\{rank(B), rank(C)\}$. Since A^{Tr} and A have the same **rank** it then follows that $rank(A) = n$. Since A has n columns the sequence of columns of A must be **linearly independent**.

Conversely, assume that the sequence of columns of A is **linearly independent**. This implies that the **null space** of A consists of the only the zero vector, $null(A) = \{\mathbf{0}_n\}$. We will show that $null(A^{Tr}A) = \{\mathbf{0}_n\}$ from which it will follow that $A^{Tr}A$ is **invertible**.

Suppose to the contrary that $null(A^{Tr}A) \neq \{\mathbf{0}_n\}$ and that $\boldsymbol{x} \in null(A^{Tr}A)$. Then $A^{Tr}A\boldsymbol{x} = \mathbf{0}_n$. Clearly the **dot product** of $\boldsymbol{x}$ and $\mathbf{0}_n$ is zero, whence

$$\boldsymbol{x}^{Tr}(A^{Tr}A\boldsymbol{x}) = 0 \tag{8.53}$$

However, $\boldsymbol{x}^{Tr}(A^{Tr}A\boldsymbol{x}) = (\boldsymbol{x}^{Tr}A^{Tr})(A\boldsymbol{x}) = (A\boldsymbol{x})^{Tr}(A\boldsymbol{x}) = (A\boldsymbol{x}) \cdot (A\boldsymbol{x})$. By the **positive definiteness** of the **dot product** it must be the case that $A\boldsymbol{x} = \mathbf{0}_m$ and so $\boldsymbol{x} \in null(A)$. However, this contradicts the assumption that $null(A) = \{\mathbf{0}_n\}$ and completes the theorem. □

Example 8.6.1. Find all the **least square solutions** for the **inconsistent linear system** $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 14 \\ 22 \\ 6 \\ 7 \end{pmatrix}$$

$$A^{Tr}A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 6 \end{pmatrix}$$

$$A^{Tr}\boldsymbol{b} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 14 \\ 22 \\ 6 \\ 7 \end{pmatrix} = \begin{pmatrix} 27 \\ 42 \end{pmatrix}$$

The matrix $A^{Tr}A$ is **invertible** so there is a unique solution

$$\frac{1}{9}\begin{pmatrix} 6 & -3 \\ -3 & 3 \end{pmatrix} \begin{pmatrix} 27 \\ 42 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$$

Fitting a Line to Data with Least Squares

It is a common feature of all scientific domains to collect data among variables and then to find a functional relationship amongst the variables that best fits the data. In the simplest case one uses a linear function. Geometrically, this amounts to finding the line which best fits the data points when graphed in a coordinate plane.

More specifically, suppose we want to fit the experimentally obtained n data points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ by a linear function $y = f(x) = a + bx$.

If the points were all collinear and on the graph of this linear function then all equations

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \\ &\vdots \\ y_n &= a + bx_n \end{aligned}$$

would be satisfied. This can be written as a matrix equation

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{8.54}$$

If we let $A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$, $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ and $\boldsymbol{m} = \begin{pmatrix} a \\ b \end{pmatrix}$ then (8.54) can be written as

$$\boldsymbol{y} = A \begin{pmatrix} a \\ b \end{pmatrix} \tag{8.55}$$

If the data points are not collinear then there will be no a and b satisfying these equations and the system represented by (8.55) is inconsistent. In this situation approximating y_i by $y'_i = a + bx_i$ results in an error $e_i = y_i - y'_i$. Now set $\boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 - y'_1 \\ y_2 - y'_2 \\ \vdots \\ y_n - y'_n \end{pmatrix}$. Now the equations become

$$\boldsymbol{y} = A \begin{pmatrix} a \\ b \end{pmatrix} + \boldsymbol{e}, \boldsymbol{y} - A \begin{pmatrix} a \\ b \end{pmatrix} = \boldsymbol{e} \tag{8.56}$$

The **least squares solution** determines the a and b such that $\parallel \boldsymbol{e} \parallel$ is minimized and is given by

$$\begin{pmatrix} a \\ b \end{pmatrix} = (A^{Tr}A)^{-1}(A^{Tr}\boldsymbol{y})$$

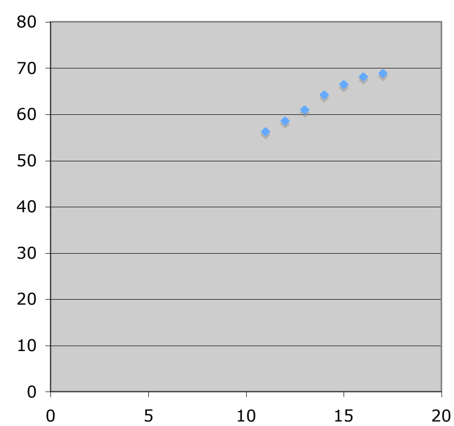
The line given by the **least squares solution** is called the ***line of best fit*** or the ***regression line***. The **norm** of the error vector $\boldsymbol{e}$ is the ***least squares error***.

Example 8.6.2. A large sample of boys, ages 11 - 17 were measured and the heights recorded. The average heights by age group are given in the following table:

Age (years)	Height (inches)
11	56.3
12	58.6
13	61.0
14	64.2
15	66.5
16	68.1
17	68.9

The graph of this data is shown in **Figure** (8.6.1). We will find the regression line of this data.

Figure 8.6.1: Plot of Age-Height Data



Let $A = \begin{pmatrix} 1 & 11 \\ 1 & 12 \\ 1 & 13 \\ 1 & 14 \\ 1 & 15 \\ 1 & 16 \\ 1 & 17 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} 56.3 \\ 58.6 \\ 61.0 \\ 64.2 \\ 66.5 \\ 68.1 \\ 68.9 \end{pmatrix}$. Then $A^{Tr}A = \begin{pmatrix} 7 & 98 \\ 98 & 1400 \end{pmatrix}$ and

$A^{Tr}\boldsymbol{y} = \begin{pmatrix} 443.6 \\ 6272.7 \end{pmatrix}$

The **reduced echelon form** of $\begin{pmatrix} 7 & 98 & | & 443.6 \\ 98 & 1400 & | & 6272.7 \end{pmatrix}$ is

$$\left(\begin{array}{cc|c} 1 & 0 & 32.22 \\ 0 & 1 & 2.23 \end{array}\right)$$

Therefore the **regression line** has equation $y = 32.22+2.23x$. The **least square error** is 1.79.

Fitting a Data to a Polynomial

Suppose you hypothesize that a set of data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ is modeled by a $k-$degree polynomial $y = f(x) = a_0 + a_1x + \cdots + a_kx^k$. We have already seen that there is a unique polynomial of degree $n - 1$ whose graph contains all the points so we may assume that $k < n - 1$. If the data points were all on the graph of this polynomial then for each i the equation

$$y_i = f(x_i) = a_0 + a_1x_i + a_2x_i^2 + \cdots + a_kx_i^k \tag{8.57}$$

would be satisfied.

Set $A = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^k \\ 1 & x_2 & x_2^2 & \ldots & x_2^k \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^k \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \boldsymbol{m} = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix}$

The equations (8.57) can be represented by the single matrix equation

$$\boldsymbol{y} = A\boldsymbol{m} \tag{8.58}$$

If the points do not all lie on some polynomial of degree at most k then the system will have no solutions $\boldsymbol{m}$. In this case, we find a best fit using the **least squares method**.

Note that the matrix obtained by taking the first $k + 1$ rows of A is

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^k \\ 1 & x_2 & x_2^2 & \ldots & x_2^k \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{k+1} & x_{k+1}^2 & \ldots & x_{k+1}^k \end{pmatrix}$$

which is a **Vandermonde matrix**. This matrix has determinant $\Pi_{i<j}(x_j - x_i) \neq 0$. Therefore, the **rank** of A is $k + 1$ and the **least squares solution** is unique and equal to

$$(A^{Tr}A)^{-1}(A^{Tr}\boldsymbol{y}) \tag{8.59}$$

We illustrate with some examples.

Example 8.6.3. Find the quadratic polynomial which is the best fit to the five points (1, -2), (2,0.2), (3,3.9), (4,10), (5, 17.9).

Set $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{pmatrix}$, $\boldsymbol{y} = \begin{pmatrix} -2 \\ 0.2 \\ 3.9 \\ 10 \\ 17.9 \end{pmatrix}$

Then $f(x) = a_0 + a_1x + a_2x^2$ where

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = (A^{Tr}A)^{-1}(A^{Tr}\boldsymbol{y}).$$

$A^{Tr}A = \begin{pmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{pmatrix}$, $A^{Tr}\boldsymbol{y} = \begin{pmatrix} 30 \\ 139.6 \\ 641.4 \end{pmatrix}$.

The **reduced echelon form** of the matrix $\begin{pmatrix} 5 & 15 & 55 & | & 30 \\ 15 & 55 & 225 & | & 139.6 \\ 55 & 225 & 979 & | & 641.4 \end{pmatrix}$ is

$\begin{pmatrix} 1 & 0 & 0 & | & -1.98 \\ 0 & 1 & 0 & | & -0.95\ 0 & 0 & 1 & | & 0.99 \end{pmatrix}$.

Therefore the quadratic polynomial which best fits these five points is

$$f(x) = --1.98 - 0.95x + 0.99x^2$$

Using this quadratic we compute the vector $\boldsymbol{y}' = \begin{pmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \\ f(5) \end{pmatrix} = \begin{pmatrix} -1.95 \\ 0.05 \\ 4.03 \\ 9.97 \\ 17.89 \end{pmatrix}$. The error vector is $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{y}' = \begin{pmatrix} -0.05 \\ 0.15 \\ -0.13 \\ 0.03 \\ 0.01 \end{pmatrix}$. The least square error is $\parallel \boldsymbol{e} \parallel = 0.04$.

Example 8.6.4. Find the cubic polynomial which is the best fit to the five points (-2,-5), (-1, 1), (0,-1), (1,-1), (2,6).

Set $A = \begin{pmatrix} 1 & -2 & 4 & -8 \\ 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{pmatrix}$, $\boldsymbol{y} = \begin{pmatrix} -5 \\ 1 \\ -1 \\ -1 \\ 6 \end{pmatrix}$.

If $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ is the cubic polynomial of best fit then $\boldsymbol{m} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}$ is the solution to $(A^{Tr}A)\boldsymbol{m} = (A^{Tr}\boldsymbol{y})$.

Direct computation gives $A^{Tr}A = \begin{pmatrix} 5 & 0 & 10 & 0 \\ 0 & 10 & 0 & 34 \\ 10 & 0 & 34 & 0 \\ 0 & 34 & 0 & 130 \end{pmatrix}$, $A^{Tr}\boldsymbol{y} = \begin{pmatrix} 0 \\ 20 \\ 4 \\ 87 \end{pmatrix}$.

The **reduced echelon form** of $\begin{pmatrix} 5 & 0 & 10 & 0 & | & 0 \\ 0 & 10 & 0 & 34 & | & 20 \\ 10 & 0 & 34 & 0 & | & 4 \\ 0 & 34 & 0 & 130 & | & 87 \end{pmatrix}$ is the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & -0.57 \\ 0 & 1 & 0 & 0 & | & -2.25 \\ 0 & 0 & 1 & 0 & | & 0.29 \\ 0 & 0 & 0 & 1 & | & 1.25 \end{pmatrix}.$$

Therefore the cubic polynomial which is the best fit to this data is

$$g(x) = --0.57 - 2.25x + 0.29x^2 + 1.25x^2.$$

Using this cubic we compute the vector $\boldsymbol{y}' = \begin{pmatrix} g(-2) \\ g(-1) \\ g(0) \\ g(1) \\ g(2) \end{pmatrix} = \begin{pmatrix} -4.93 \\ 0.71 \\ -0.57 \\ -1.29 \\ 6.07 \end{pmatrix}$. The error vector is $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{y}' = \begin{pmatrix} -0.07 \\ 0.29 \\ -0.43 \\ 0.29 \\ -0.07 \end{pmatrix}$. The least square error is $\parallel \boldsymbol{e} \parallel = 0.60$.

Fitting Data to an Exponential Function

Sometimes the graph of some data or the context in which it is collected suggests that the most appropriate approximation for the data is by an exponential function; for example, growth of the national income or the amount of a radioactive material present at given time intervals.

Thus, given some points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ we wish to approximate this data by a function $y = Ce^{kt}$ for some constants C and k.

Note that for such a function, $ln\, y = ln\, C + kt$ is a linear function of t. We can therefore use the **least squares method** for finding $ln\, C$ and k from the data $(x_1, ln\, y_1), (x_2, ln\, y_2)$, $\ldots, (x_n, ln\, y_n)$.

Example 8.6.5. Find the exponential function $y = Ce^{kt}$ which best approximates the following 6 data points:

$$(-2, .14), (-1, .32), (0, .55), (1, 1.24), (2, 2.44), (3, 4.75)$$

Taking the natural logs of the $y-$ values we get the data points:

$$(-2, -1.97), (-1, -1.14), (0, -.60), (1, .22), (2, .89), (3, 1.56)$$

We now need to find the **least squares solution** to $A\boldsymbol{x} = \boldsymbol{b}$ where $A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} -1.97 \\ -1.14 \\ -.60 \\ .22 \\ .89 \\ 1.56 \end{pmatrix}$. The matrix form of the **normal equations** for this **least squares problem** is as follows:

$$A^{Tr}A\boldsymbol{x}' = A^{Tr}\boldsymbol{b}, \begin{pmatrix} 6 & 3 \\ 3 & 19 \end{pmatrix} \boldsymbol{x}' = \begin{pmatrix} 1.04 \\ 11.76 \end{pmatrix}$$

The solution to this is $\boldsymbol{x}' = \begin{pmatrix} -.524 \\ .702 \end{pmatrix}$. Then $C = e^{-.524} = .59, k = .702$. Since $e^{.702} \sim 2.02$ the data is approximated by the function $h(x) = .59(2.02)^t$. We compute the vector $\boldsymbol{y}' = \begin{pmatrix} h(-2) \\ h(-1) \\ h(0) \\ h(1) \\ h(2) \\ h(3) \end{pmatrix} = \begin{pmatrix} 0.14 \\ 0.29 \\ 0.59 \\ 1.19 \\ 2.41 \\ 4.86 \end{pmatrix}$. The error vector is $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{y}' = \begin{pmatrix} 0 \\ 0.03 \\ -0.04 \\ 0.05 \\ 0.03 \\ -0.11 \end{pmatrix}$.

The least square error is $\parallel \boldsymbol{e} \parallel = 0.13$.

The QR computation of Least Squares Solutions

Once again consider the **least squares problem** to an **inconsistent linear system** equivalent to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ where A is an $m \times n$ matrix and $\boldsymbol{b} \in \mathbb{R}^m$. In some instances the coefficients in the **normal equations** are highly sensitive to small changes, that is, small errors in the calculation of the entries in $A^{Tr}A$ can cause significant errors in the solution of $\boldsymbol{x}'$. When the matrix $A^{Tr}A$ is **invertible** it is therefore sometimes better to calculate the **least squares solution** using the **QR factorization** of the matrix A.

We recall that if A is an $m \times n$ matrix then there is an $m \times n$ matrix Q whose columns are an **orthonormal sequence** (and a **basis** for $col(A)$) and an **invertible** $n \times n$ **upper triangular matrix** R such that $A = QR$.

In this case the matrix form of **normal equations**, $(A^{Tr}A)\boldsymbol{x}' = A^{Tr}\boldsymbol{b}$, becomes

$$[(QR)^{Tr}(QR)]\boldsymbol{x}' = (QR)^{Tr}\boldsymbol{b} \tag{8.60}$$

Using the fact that $(BC)^{Tr} = C^{Tr}B^{Tr}$ (8.60) becomes

$$[R^{Tr}(Q^{Tr}Q)R]\boldsymbol{x}' = R^{Tr}Q^{Tr}\boldsymbol{b} \tag{8.61}$$

Since Q is an **orthonormal matrix** $Q^{Tr}Q = I_n$. Also, since R is **invertible**, so is R^{Tr} and therefore it can be canceled from both sides. Making use of these two conditions (8.61) now becomes

$$R\boldsymbol{x}' = Q^{Tr}\boldsymbol{b}, \boldsymbol{x}' = R^{-1}(Q^{Tr}\boldsymbol{b}) \tag{8.62}$$

Example 8.6.6. Find the **least squares solution** to the **inconsistent system** $A\boldsymbol{x} = \boldsymbol{b}$ where $A = \begin{pmatrix} 1 & 3 & 4 \\ 1 & 3 & 2 \\ 1 & -1 & 2 \\ 1 & -1 & 0 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 2 \\ 8 \\ 4 \\ 6 \end{pmatrix}$.

The **Gram-Schmidt process** yields the following **orthonormal basis** for $col(A)$:

$$\{\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}\}.$$

Set $Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$. Q is an **orthonormal matrix** and $col(Q) = col(A)$. If we set $R = Q^{Tr}A = \begin{pmatrix} 2 & 2 & 4 \\ 0 & 4 & 2 \\ 0 & 0 & 2 \end{pmatrix}$ then $A = QR$. $R^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} & -\frac{3}{4} \\ 0 & \frac{1}{4} & -\frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}$, $Q^{Tr}\boldsymbol{b} = \begin{pmatrix} 10 \\ 0 \\ -4 \end{pmatrix}$ and $\boldsymbol{x}' = R^{-1}(Q^{Tr}\boldsymbol{b}) = \begin{pmatrix} 8 \\ 1 \\ -2 \end{pmatrix}$.

Exercises

In exercises 1 - 4 show that the given vector $\boldsymbol{b}$ is not in the **column space** of the given matrix A. Verify that the columns of A are **linearly independent**. Write down the **normal equations** for the **least squares solution** to the **linear system** $A\boldsymbol{x} = \boldsymbol{b}$ and find the unique **least square solution** $\boldsymbol{x}'$.

1. $A = \begin{pmatrix} 1 & 1 \\ 1 & -3 \\ -2 & 2 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} 9 \\ 3 \\ -6 \end{pmatrix}$

2. $A = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 3 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} 1 \\ -2 \\ 7 \end{pmatrix}$

3. $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \\ 18 \end{pmatrix}$

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

In exercises 5 and 6 show that the given vector $\boldsymbol{b}$ is not in the **column space** of the given matrix A. Verify that the columns of A are **linearly dependent**. Write down the **normal equations** for the **least squares solution** to the **linear system** $A\boldsymbol{x} = \boldsymbol{b}$ and find the **general least square solution** $\boldsymbol{x}'$.

5. $A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 3 & -1 \\ 2 & 3 & 4 \\ 3 & 6 & 3 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ 0 \\ 4 \\ 2 \end{pmatrix}.$

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

In exercises 7 and 8 verify that the given **orthonormal sequence** is a **basis** for the **column space** of the given matrix A. Use this to obtain a **QR factorization** for A and apply this to find the **least square solution** to the **inconsistent linear system** $A\boldsymbol{x} = \boldsymbol{b}$ for the given vector $\boldsymbol{b}$.

7. $A = \begin{pmatrix} 1 & 1 \\ 2 & 8 \\ -2 & -5 \end{pmatrix}, \mathcal{O} = \{ \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \\ -\frac{2}{3} \end{pmatrix}, \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix} \}, \boldsymbol{b} = \begin{pmatrix} 2 \\ 7 \\ 5 \end{pmatrix}.$

8. $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 2 \\ 1 & -3 & -1 \\ 1 & -3 & -2 \end{pmatrix}, \mathcal{O} = \{ \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \}, \boldsymbol{b} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 4 \end{pmatrix}$

In exercises 9 and 10 find the **regression line** and the **least squares error** for the given data.

9. (-2, -3.8), (-1, -1.1), (0, 1.9), (1, 5.2), (2, 8.1)

10. (-1, 3.3), (0,1.6), (1, -.8), (2, -2.5), (3, -4.4)

Table 8.1: Average Hourly Wages, 1990 - 2003, in Constant Dollars

Year	Wages
1990	7.66
2000	8.03
2001	8.11
2002	8.24
2003	8.27

11. Average hourly wages of American workers for the several years from 1990 to 2003 are given in Table (8.1) . Find the **line of best fit** for this data.

In exercises 12 and 13 find the quadratic polynomial which best approximates the given data. See **Fitting Data to a Polynomial Function** and **Example** (8.6.3).

12. (-1, 4.1), (0, 2.3), (1, 2.6), (2, 4.2), (3, 8.2)

13. (-1, 1.0), (0, .7), (1, 1.2), (2, 2.5), (3, 5.0), (4, 8.7)

In exercises 14 and 15 find the exponential function $y = Ce^k t$ which best approximates the given data. See **Fitting Data to an Exponential Function** and **Example** (8.6.5).

14. (0, .4), (1, .9), (2, 1.3), (3, 2.5), (4, 3.3)

15. (-2, 3.1), (-1, 2.8), (1, 2.3), (2, 2.0), (4, 1.6)

Table 8.2: Consumer Price Index, 1990 - 2003

Year	CPI
1990	131
1995	152
1999	167
2000	172
2001	177
2002	180
2003	184

16. Table (8.2) shows the Consumer Price Index for certain years between 1990 and 2003.

a) Find the **least squares approximating line** for this data and the **least square error** for this line. Also determine the projected CPI in 2010 for this model. See **Fitting a Line to Data with Least Squares** and **Example** (8.6.2).

b) Find the **least squares approximating exponential** for this data and the corresponding least squares error. Also determine the projected CPI in 2010 for this model. See **Fitting Data to an Exponential Function** and **Example** (8.6.5).

c) Which is a better appproximation. Explain your answer.

Table 8.3: US Imports of Crude Oil in Million Metric Tons, 1990 - 2003

Year	Imports
1990	2151
1995	2639
1997	3002
1998	3178
1999	3187
2000	3311
2001	3405
2002	3336
2003	3521

17. Table (8.3) shows US imports of crude oil in millions of metric tons for years 1990 to 2003.

a) Find the **least squares approximating line** for this data and the **least square error** for this line. Also determine the projected imports of crude oil for the year 2050. See **Fitting a Line to Data with Least Squares** and **Example** (8.6.2).

b) Find the **least squares approximating quadratic** for this data and the corresponding least squares error. Also determine the projected imports of crude oil in 2050 for this model. See **Fitting Data to a Polynomial Function** and **Example** (8.6.3).

c) Find the **least squares approximating exponential** for this data and the corresponding least squares error. Also determine the projected crude oil imports in 2050 for this model. See **Fitting Data to an Exponential Function** and **Example** (8.6.5).

d) Which is the best approximation to the given data? Explain your answer.

Table 8.4: Average CEO Total Compensation (Wages and Options), 1992-2000

Year	Total Compensation (In Million $)
1992	3.5
1993	3.3
1994	4.0
1995	4.4
1996	6.1
1997	7.9
1998	10.4
1999	10.9
2000	14.8

18. Table (8.4) shows average CEO compensation for the Fortune 800 for the years 1992 - 2000.

a) Find the **least squares approximating quadratic** for this data and the corresponding least squares error. Also determine the projected average CEO compensation in 2015. See **Fitting a Line to Data with Least Squares** and **Example** (8.6.2).

b) Find the **least squares approximating exponential** for this data and the corresponding least squares error. Also determine the projected average CEO compensation in 2015. See **Fitting Data to an Exponential Function** and **Example** (8.6.5).

c) Which is the best approximation to the given data? Explain your answer.