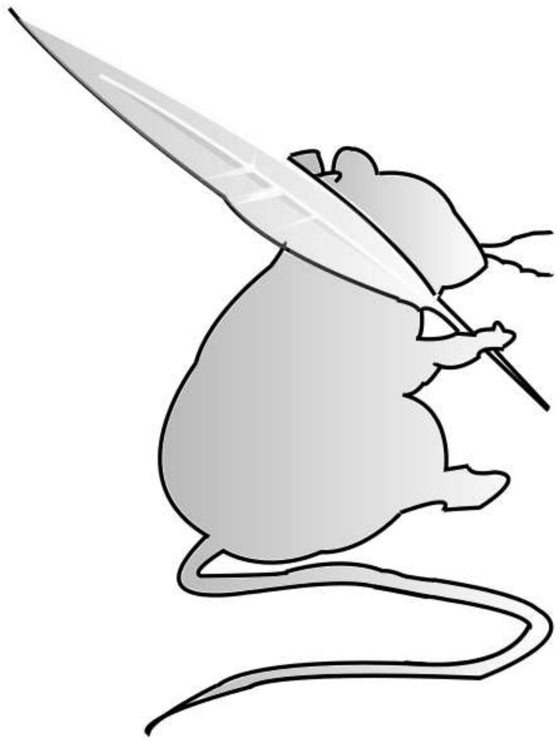


Linear Algebra
for Mouse, Pen and Pad

February 9, 2026

Preface

This text is intended to be suitable as an introduction to linear algebra for senior high school and community college students and especially for those people who desire to learn about the subject but have been away from formal education for a while. Knowledge of calculus and analytic geometry is not assumed. It is expected that the reader has some familiarity with Euclidean geometry.[1]

Linear algebra is easily defined as the study of **linear transformations**, **matrices** and **vector spaces**. This text and its accompanying software package, called **Lampp**, were designed to help students understand what linear transformations and vectors spaces are, as these concepts are not so casually definable. Introductory applications are given for physics, chemistry, economics, etc., but greater attention is paid to proof as the intention is to allow accommodation without compromise of rigor. The objective is to give a solid foundation leading gently, but rapidly, to mathematically sophisticated concepts.

The exercises in this text are divided into two sections, the mandatory (featuring proofs, derivations and *problems*, which must be accomplished with pen, paper and thought) and the optional (composed of application *exercises* which can be facilitated with the use of a computer). Answers to selected problems and exercises are given in separate sections near the end of the text.

The text of this book stands alone but is best used in conjunction with Lampp, an app used for working with numerical examples. The Lampp software, written in JavaFX, can be run as a standalone application on many kinds of tablets or computers. Lampp also provides a convenient portal to the textbook's website where more dynamic demonstrations and further graphic examples of key concepts are made available.

Fields and number systems are not traditionally taught until university. It is hoped that an earlier acquaintance with these concepts will allow teachers to help students view number systems other than the **reals** in an intuitive manner. The experience and familiarity gained should naturally motivate a desire for further understanding.

The traditional cross product, limited to three dimensions, is shown but is subsumed by the introduction of the outer product and the geometric product, which are not so limited. This was done to introduce the student to **geometric**

[1] If one feels a need for a good review, I would suggest Chrystal's public domain text on algebra (see [5] in the References) and Slaught and Lennes' public domain book on geometry (see [17].)

algebra. Geometric algebra, which builds on concepts of geometry and linear algebra, is a modern approach to mathematical reasoning based on the works of Hermann Grassmann (1809-1877) and William Kingdon Clifford (1845-1879). David Hestenes is the foremost proponent of this mathematical tool and deserves much praise. Geometric algebra, restricted to real numbers, is exceedingly useful for modeling coordinate-free geometry and physical applications. It is very unlikely that anyone involved in a future study of linear algebra, differential geometry, projective geometry, vector calculus, tensor analysis, Lie algebra, algebraic geometry, quaternions, complex analysis, 3D modeling or advanced physics will *not* become acquainted with geometric algebra.

While writing this text, the author stumbled upon Mamikon's Sweeping-Tangent Theorem and has introduced some of Mamikon A. Mnatsakanian's wonderful discoveries as they tie in with geometric algebra and linear algebra. The use of tangent sweeps and tangent clusters rapidly leads to visualization of the equivalence of specific subsets of vector spaces. This naturally leads to an appreciation of the enormous breadth of integral calculus (and some of its limitations!)

The use of set theory in presentation is initially minimized in this text. Set notation and set theory concepts are introduced as needed.

The use of Lampp with this text:

Carl Friedrich Gauss (1777-1855) was said to recall tables of logarithms because it was easier (for him!) than getting out of his chair to walk across the room to get a book. While this is usually held as an example of his remarkable mental prowess, it also seems to indicate that even Gauss would have used a reference book if it had been closer to hand. It is hoped that the use of Lampp will be like bringing the book[1] closer to hand, freeing the student from mundane, error-prone calculations while allowing them to concentrate on familiarizing themselves with the power of the fundamental concepts of linear algebra.

There are many very good programs already available for solving systems of equations. However, few were designed strictly as teaching aids and most emulate the rational number field with the use of floating point numbers, which is an extra level of abstraction which can be avoided.

Lampp (Linear Algebra Matrix aPP)[2] was written in such a way that it is impossible to use correctly if one does not understand what fields and row operations are. The user begins by first choosing a predefined field. A matrix of up to 10 by 10 scalars may then be created and the user then selects which row (or column) operations they wish to perform upon the matrix. Multiplication and addition of field scalars are then facilitated. Predefined fields include rational numbers, complex with rational coefficients (called Gaussian) and integers modulo a selected prime. Matrix multiplication and addition are accomplished by showing

[1] Paul Erdős' book: the metaphysical book of mathematics containing only the most elegant mathematical proofs.

[2] named to coincide with *Linear Algebra for Mouse, Pen and Pad*

every intermediate step so that a student (or instructor) may check a problem worked out by hand.

As the student works their way through the text, keys become available to unlock further capabilities of Lampp. For example, when the student is at the point where there is a need to easily create the transform of a matrix, a hidden feature in Lampp is shown so that simple operations on matrices may be performed without having to accomplish every intermediate step.

For those who wish to modify Lampp (by adding other fields, for instance) the JavaFX source code is made available from GitHub. The source code is distributed following the GNU copyleft convention. More information about this is given in the appendix *Lamp on your Computer.* Lampp has been tested extensively with the exercises in this book but comes with no guarantee.

Contents

Introduction - The subject of linear algebra

Learning to do arithmetic was a great accomplishment for mankind as well as for every person who struggled with it. Adding, subtracting, learning multiplication tables and applying this hard-won knowledge to division were milestones in personal achievement. Certainly very few adults regret the efforts that they were encouraged to put forth to attain these skills as children. The utility of arithmetic is evident even in an age (maybe even especially in an age) of hand-held computers and user-friendly software packages.

The next great milestone was the concept of representation of arbitrary numbers by symbols. We learn to use symbols like a, b, c, y, z and, of course, the ubiquitous x, to stand for unknown quantities in arithmetical expressions. Instead of $1 + 1 = 2$, we now have $a + 1 = 2$, so what is a? We refer to these unknown quantities as **variables**. If a symbol refers to a fixed number, we refer to that symbol as a **constant**. Usually constants are written as symbols because it is difficult or impossible to write out a number's exact expression.

This is the discovery of **algebra**. It is the use of letters and symbols to stand for numbers in arithmetical and other mathematical statements and the rules used to manipulate these numbers.

Another great invention is **geometry**. In this study we learn to manipulate the concepts of **points** and **lines**. When we merge the ideas of geometry with algebra we are proceeding into an area known as **analytic geometry**. A review of some concepts of analytic geometry used in this text will be given in the first chapter.

A Brief History of Linear Algebra

Although many of the mathematical ideas which are embraced by linear algebra have been used by mathematicians for centuries the name itself is fairly recent. Most of the development has taken place since the mid 1800s from attempts to discover general methods to solve systems of linear equations.

Curiously most efforts were originally centered around the study of **determinants**. This concept (which usually isn't encountered until well into most modern studies of linear algebra) was developed by Gottfried Wilhelm Leibniz in 1693. Gabriel Cramer further developed this notion, along with Joseph-Louis Lagrange, Carl Friedrich Gauss and others.

In 1850, James Joseph Sylvester used the term **matrix** to designate a rectangular array (table) of numbers. These were used to calculate determinants. Sir Arthur Cayley, in 1855, made discoveries involving matrix algebra and made the very prophetic remark:

> There would be many things to say about this theory of matrices which should, it seems to me, precede the theory of determinants.

At nearly the same time, William Rowan Hamilton also used the algebra of matrices to study linear and **vector** functions. Vectors, which often are represented as ordered lists of numbers, were first used to define physical quantities (like force and velocity) which had magnitude and direction. One needed two numbers to represent such values. For example, a velocity vector written as $(12, \pi)$ could mean that we are going 12 kilometers per hour in direction π (which usually means directly to the left. This is an example of **radian** coordinates which will be introduced in a later chapter.)

During this same period, a little known German schoolteacher named Hermann Gunther Grassmann also did much work in the field of linear algebra, proving many theorems (one of which was named after a mathematician who published a paper on his rediscovery 26 years after Grassmann's death, see [7]). Grassmann's work on the connections between algebra and geometry are only now being generally recognized for their beauty and utility.

Linear algebra is still very much a lively area of research and its uses in other types of mathematics, and in economics, social sciences, physics, chemistry and computer science, expand and grow daily.

Chapter 1

Matrices and Analytic Geometry

1.1 Some Mathematical Notation

The symbols used by mathematicians (the notation) are very important for clear presentation. A good notation even can lead to the development of ideas. Perhaps the most famous example of this is the notation invented by Leibniz for calculus which eventually displaced the English form developed by Newton.

People who are interested in learning mathematics often have difficulty with the notation. This difficulty can usually be erased if they keep in mind that new symbols are used to save reading time and writing space. An example is given below.

Suppose we wished to define a mathematical formula to express the total price of several computer systems. Using algebra we would let $x =$ *the cost of one computer system*. Since x is an arbitrary cost we use **subscripts** to refer to individual computer systems. If we had four computer systems we could write a numbered list (which we say is a kind of **set**):

$x_1 =$ *the cost of the first system*
$x_2 =$ *the cost of the second system*
$x_3 =$ *the cost of the third system*
$x_4 =$ *the cost of the fourth system*

To be even more abstract we could say that $x_i =$ *the cost of the* i^{th} *computer system*, for $i = 1, 2, 3$ or 4. In spoken English, we refer to x_i as *x sub i*. We also refer to each x_i as an **entry** of a list or **ordered set**.

We would say that the total cost of these four computer systems is

$$x_1 + x_2 + x_3 + x_4 = x_{total}$$

To take advantage of the benefits of mathematical notation we write:

$$\sum_{i=1}^{4} x_i = x_{total}$$

In spoken form we might refer to this as *sum with i from 1 to 4 of x sub i equals x sub total.*

We call i an **index**. Its value gives us the position of an entry. When it is clear what they are talking about some authors omit the index and its range from the sum notation $\sum$. They would use

$$\sum x \text{ to mean } \sum_{i=1}^{n} x_i$$

Here n is the number of values of x.

In our example above, we let i range from 1 to 4. A range starting at 1 is what we will typically use in this text and is most common.[1] The total cost of two computers, the second and the third, would be written

$$\sum_{i=2}^{3} x_i = x_2 + x_3$$

When faced with a new symbol in a mathematics text the student is advised to write out several examples of the meaning using notation already known. When one becomes really irked with writing out the same thing over and over one can then appreciate the reason for adopting the use of the new symbol. Notice that this is the same kind of reasoning that keeps us from hitting our thumbs with hammers.

The use of subscripts actually saves a lot of writing, especially when we deal with **matrices**. As was mentioned in the preface, a matrix is a rectangular arrangement of numbers or symbols (which usually, but not always, stand for numbers). An example of a matrix is given below.

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

Subscripts are used to refer to the **entry**. Examples are

$$m_{1,2} = 2$$

$$m_{2,2} = 4$$

The first subscript of $m_{1,2}$ refers to the first **row** of the matrix M. The second subscript, 2, refers to the second **column** of the matrix. Horizontal lists of numbers are rows and vertical lists of numbers are columns.

[1] Some computer languages start the index at 0, so a first item would be called x_0. The authors of the languages did this to take advantage of **modular arithmetic**. We will also make use of modular arithmetic later on.

Entries may also be many other things. Valid entries could be a function that return a number or even a matrix itself. We show two possible entries for matrices A and B below. We show that an entry could even be another matrix.

$$a_{3,2} = \sqrt{x + 2y}$$

$$b_{1,3} = \begin{pmatrix} 5 & 2 \\ 7 & 8 \end{pmatrix}$$

$$b_{4,4} = \begin{pmatrix} a & 2b \\ 3c & 5 \end{pmatrix}$$

You may have noticed that we used lowercase m to refer to an entry of a matrix and the uppercase M to refer to the matrix itself. In this text we will use capital letters to name matrices, as we also did when we showed entries for A and B.

Also notice the difference between subscripts and **superscripts**. For example, in the expression x^2, 2 is a superscript. Superscripts are used to indicate **powers** of numbers (the number of times we multiply a number by itself). For instance, $x^2 = x * x$, $x^3 = x * x * x$ and $y_2^4 = y_2 * y_2 * y_2 * y_2$.

We can use the sum notation to refer to entries in a matrix. For example:

$$\sum_{j=1}^{2} m_{2,j}$$

would give us the total of the two entries in the second row of our matrix M, which would be 7.

The expression

$$\sum_{j=1}^{2} \sum_{i=1}^{2} m_{i,j}$$

would give us the total of all the entries in the matrix M. This can be seen by expanding the innermost sum sign:

$$\sum_{j=1}^{2} m_{1,j} + m_{2,j}$$

and then expanding the last sum sign:

$$m_{1,1} + m_{2,1} + m_{1,2} + m_{2,2}$$

You should notice that if we had exchanged the order of the two sum signs, we would still get the same answer.

$$\sum_{i=1}^{2} \sum_{j=1}^{2} m_{i,j}$$

would expand to

$$m_{1,1} + m_{1,2} + m_{2,1} + m_{2,2}$$

which totals to the same value. (This is because the order in which we add terms doesn't matter for ordinary numbers. We will have more to say about this later.)

A general matrix X with m rows and n columns is usually written as

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \ldots & x_{1,n} \\ x_{2,1} & x_{2,2} & x_{2,3} & \ldots & x_{2,n} \\ x_{3,1} & x_{3,2} & x_{3,3} & \ldots & x_{3,n} \\ \vdots & & & & \vdots \\ x_{m,1} & x_{m,2} & x_{m,3} & \ldots & x_{m,n} \end{pmatrix}$$

We say that matrix X is m by n or has **order** m by n. We would write this as $X_{m\times n}$. A matrix Q that had 2 rows and 4 columns would be denoted as $Q_{2\times 4}$.

A matrix X and a matrix W are said to be **equal** if they are of the same order and $x_{i,j} = w_{i,j}$.

In order to save ink and pixels, if a matrix has just one row or one column, only one subscript will be used. Some examples are

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \text{ and } Z = \begin{pmatrix} z_1 & z_2 & z_3 & z_4 \end{pmatrix}$$

Furthermore, any matrix $X_{1\times 1}$ (any matrix with only one entry) will simply be referred to as another number x without any subscripts.

Exercises 1.1

1. For the matrix
$$X = \begin{pmatrix} 4 & 7 & 12 \\ 3 & 10 & 1 \end{pmatrix}$$
$$\text{What is } \sum_{j=1}^{3} x_{1,j} \text{ ? What is } \sum_{j=1}^{3} x_{2,j} \text{ ?}$$

2. How would you write the sum notation for the total value of all the entries of the matrix X in the previous exercise?

3. What is the sum notation for the total value of all the entries of the matrix $Y = \begin{pmatrix} 5 & 6 & 3 & 2 \end{pmatrix}$?

4. What would be the value of $\sum_{i=1}^{4} i$?

5. What is the value of $\sum_{k=1}^{3} \sum_{j=1}^{2} k + j$?

6. Show that $\sum_{i=1}^{3} \sum_{j=1}^{i} j = 11$ by writing out each step.

7. Show that $\sum_{i=1}^{3} \sum_{j=i}^{3} j = 14$.

1.2 Number Systems

Most applications of mathematics involve the use of the **real** number system. We define these numbers as any number than can be expressed in **decimal notation**. It is easier to give examples of such numbers than to define them properly. For instance, $2 \quad 0.0004 \quad 3.1415926\ldots$, are three examples of real numbers. (The ... means that the number could be continued indefinitely.)

We will use the symbol $\mathbb{R}$ to mean the real number system. We will also use the symbol $\in$ to show that a number is a member of a number system. So $x \in \mathbb{R}$ would mean that x is a real number. (Read this as *x in the reals*). To show that x is **not** a real number we write $x \notin \mathbb{R}$. In reference to set notation, we further define the symbol $\in$ to mean that an object is an **element** of a set. For instance, we say $e_2 \in E$, *e sub 2 in E*, or, another way, e_2 is an element of the set E. A set is any defined collection of objects (elements), like a list (items are elements), a field (numbers are elements), etc.

There are other number systems. There is the set of **integer** numbers, $(\ldots, -2, -1, 0, 1, 2, \ldots)$, which we refer to with the symbol $\mathbb{Z}$.[2] The **positive integers**, $\mathbb{Z}^+$, are all the integers greater than 0. These are the numbers most commonly used as indices of matrices. There are the **rational** numbers formed by the ratio of an integer over a non-zero integer, like $\frac{1}{2}$ or $\frac{-23}{1}$. We refer to these numbers with the symbol $\mathbb{Q}$.

There are other number systems, some of which we will introduce later. A natural question is why do we need different number systems? The simple answer is that problems we encounter are best handled by certain types of numbers. When we count people we don't need rational numbers. When we measure liquids, like liters of juice, we almost never need to introduce real numbers that are not ratios of integer numbers (rationals).

All the number systems which we will talk about obey certain **laws** of algebra. We will examine how these laws apply to the rational numbers.

We mentioned before that a rational number is a number of the form q/r where q and r are integer numbers and $r \neq 0$ (read *r is not equal to zero*). There are two **binary compositions** that we normally use with the rational numbers. These are **addition** and **multiplication**. That is, addition and multiplication are operations where we combine two numbers (binary means two) to create a third number (composition). We will use the symbol $+$ to mean addition. We will use the symbol $*$ to mean multiplication or no symbol at all when we are using characters to represent numbers. For example, pq and $p * q$ would mean the same thing, p multiplied by q. Also $4s$ is 4 times s, but 34 is **not** $3 * 4$, it is thirty-four.

We say that two rational numbers, q/r and s/t, are equal if $qt = rs$. In other words, if q times t equals r times s. Using this definition we know then that $\frac{3}{4} = \frac{6}{8}$ because $3 * 8$ equals $4 * 6$.

A rational number is written either as p/q or $\frac{p}{q}$. Here p is called the **numerator** and q is called the **denominator**. So for $\frac{3}{4}$, 3 would be the numerator and

[2]The use of $\mathbb{Z}$ was adopted because the German word for 'number' begins with Z.

4 would be the denominator.

Addition of two rationals is accomplished using the following formula:

$$\frac{p}{q} + \frac{r}{s} = \frac{ps + qr}{qs}$$

$$\frac{2}{3} + \frac{4}{7} = \frac{2 * 7 + 3 * 4}{3 * 7} = \frac{14 + 12}{21} = \frac{26}{21}$$

Notice that the multiplication was done **before** the addition. This is the accepted **order of operations**. When doing arithmetic, multiplication is done before addition unless an operation (another word for composition) is surrounded by a pair of braces. For example:

$$(2 + 4) * ((3 + 1) + 4 * 5) = (6) * ((4) + 20) = 6 * 24 = 144$$

A clearly superior method for calculating the result of binary operations is the **postfix notation**. In postfix notation, two numbers or variables are followed by the operation symbol. An expression such as $3\ 4\ +$ would be the same as $3+4$ and $x\ y\ *$ would be xy. All binary operations, where two consecutive numbers or variables are immediately followed by an operation symbol, are done first. The preceding example would be written as

$$2\ \ 4\ \ +\ \ \ 3\ \ 1\ \ +\ \ \ 4\ \ 5\ \ *\ \ +\ \ *\ \ \ =$$

$$6\ \ 4\ \ 20\ \ +\ \ *\ \ \ =$$

$$6\ \ 24\ \ *\ \ \ =$$

$$144$$

Although postfix notation removes the need for braces and memorizing orders of operations, it is not in common use as yet so we will reluctantly do without it in this text.

The sum notation has the same order of operations as a pair of braces so that

$$5\sum_{i=1}^{3} i + 1 = 5 * (2 + 3 + 4) = 45$$

and that

$$5\left(\sum_{i=1}^{3} i\right) + 1 = 5 * (1 + 2 + 3) + 1 = 31$$

Multiplication of rationals is much simpler than addition of rationals. We simply multiply the numerators and multiply the denominators.

$$\frac{p}{q} * \frac{r}{s} = \frac{pr}{qs}$$

$$\frac{2}{3} * \frac{4}{7} = \frac{2 * 4}{3 * 7} = \frac{8}{21}$$

Rationals are usually written in their **lowest form**. This means that we **divide out** the **greatest common divisor**. For instance:

$$\frac{24}{40} = \frac{3*8}{5*8} = \frac{3}{5}$$

Since $\frac{3*8}{5*8} = \frac{3}{5} * \frac{8}{8} = \frac{3}{5} * 1$. In this example, divisors of 24 are 3 and 8, divisors of 40 are 5 and 8 and the largest common divisor of 24 and 40 is 8. A formal definition of divisor would be

Definition 1.1. *If a and c are integers and $c \neq 0$, then c is a **divisor** (or **factor**) of a if there exists an integer b such that $a = b * c$.*

The greatest common divisor (sometimes written g.c.d.) is defined as

Definition 1.2. *If $a \neq 0$ and b are integers, the **greatest common divisor** of a and b is the unique positive integer d such that d is a divisor of both a and b and if some integer c is a divisor of both a and b, then c is a divisor of d.*

The two binary compositions we use with the rationals always produce another rational number. When this is true for some number system we say that the number system is **closed** under that binary composition. (We would say that the rationals are closed under addition or that the reals are closed under multiplication.)

We have made use of laws of algebra to accomplish the above calculation. We will now state the five essential **algebraic laws for the rational numbers**.

1. **The associative laws**

   addition $(x + y) + z = x + (y + z)$ for any x,y and $z \in \mathbb{Q}$.

   multiplication $(x * y) * z = x * (y * z)$ for any x,y and $z \in \mathbb{Q}$.

2. **The commutative laws**

   addition $x + y = y + x$ for any $x, y \in \mathbb{Q}$.

   multiplication $x * y = y * x$ for any $x, y \in \mathbb{Q}$.

3. **The distributive law**

   $x * (y + z) = x * y + x * z$ for any rational numbers x, y, and z.

4. **Identity laws**

   addition There exists a unique rational number 0 such that $0 + x = x$ for any $x \in \mathbb{Q}$.

   multiplication There exists a unique rational number 1 such that $1 * x = x$ for any rational number x.

5. **Inverse laws**

addition For any rational number x, there is a unique rational number $(-x)$ such that $x + (-x) = 0$. We call $-x$ the **additive inverse** of x.

multiplication For any rational number $x \neq 0$, there is a unique rational number (x^{-1}) such that $x*(x^{-1}) = 1$. We call x^{-1} the **multiplicative inverse** of x.

Problems 1.2

1. Not all binary compositions are associative. Subtraction, for instance, is not associative. Give an example to show that subtraction of rationals is not associative.

2. Subtraction is also not commutative. Is division commutative?

3. For a rational number $x = \frac{p}{q}$, what would be x^{-1}?

4. A simpler way of writing $-(-b)$ is simply b. What is a simpler way of writing $(w^{-1})^{-1}$? Assume that $b, w \in \mathbb{Q}$.

5. The **absolute value** of a number r is $-r$ if r is less than zero or simply r if r is greater than or equal to zero. When we write $|x|$ we are asking for the absolute value of x. Is $|x + y| = |x| + |y|$ for all $x, y \in \mathbb{Q}$?

6. Give an example that shows that the integers are not closed under division.

7. Show that the rationals are not closed under division.
(**HINT**: $\frac{a}{b}/\frac{c}{d} = \frac{a}{b} * \frac{d}{c}$.)

8. When we have $y = x^2$, $|x|$ is called the **square root** of y. We write this as $\sqrt{y} = x$. Notice that, by definition, x will always be a positive number. If y is any real number, will x also be a real number?

Exercises 1.2

1. A **prime** number is a positive integer which leaves a remainder if it is divided by any integer greater than one and less than the number. Examples are $2, 3, 5, 7, 11, 13, 17, \ldots$. All other integers greater than one are called **composite** numbers. When we **factor** a composite number into primes, we give an arithmetical expression of all the prime divisors which multiply together to equal the composite number. A negative integer less than -1 is factored into primes by factoring its absolute value and then multiplying by -1. Two examples would be $15 = 3 * 5$ and $18 = 2 * 3 * 3$. Factor the following numbers into primes: $12, 30, 32, -45$.

2. What is the lowest form of $\frac{12}{30}$? Of $\frac{12}{32}$? Of $\frac{32}{30}$? Of $\frac{-45}{30}$?

3. What is $\frac{3}{4} + \frac{1}{2}$? Give your answer in the lowest form.

4. If $x = \frac{3}{7}$ and $y = \frac{3}{5}$, what is $x^{-1} * y$? What is $x * y^{-1}$?

5. If $a = \frac{2}{17}$, what is $(a^{-1})^{-1}$?

1.3 Fields

The numbers in a number system are sometimes referred to as **scalars**.

Definition 1.3. *A **field** $\mathbb{F}$ is a collection of scalars closed under two binary compositions $+$ and $*$ such that*

1. *If x, y, z are in the field $\mathbb{F}$ then $(x+y)+z = x+(y+z)$.*
2. *Given $x, y, z \in \mathbb{F}$ then we have $x*(y*z) = (x*y)*z$.*
3. *Given x, y in the field $\mathbb{F}$ then $x+y = y+x$.*
4. *If x and y are in $\mathbb{F}$ then $xy = yx$.*
5. *For any $x \in \mathbb{F}$ there exists a unique 0 such that $0+x = x+0 = x$.*
6. *The field $\mathbb{F}$ must contain a unique scalar called 1 such that $1 \neq 0$ and $x*1 = 1*x = x$ for any $x \in \mathbb{F}$.*
7. *If $x \in \mathbb{F}$ then this implies that there is a unique $(-x) \in \mathbb{F}$ such that $x+(-x) = (-x)+x = 0$.*
8. *If the three scalars x, y, z are in the field $\mathbb{F}$ then we must have that $x*(y+z) = x*y+x*z$ and $(x+y)*z = x*z+y*z$.*
9. *For any $x \neq 0$ in $\mathbb{F}$ there is a unique $x^{-1} \in \mathbb{F}$ such that $x*x^{-1} = x^{-1}*x = 1$.*

All of these properties are necessary in order for $\mathbb{F}$ to be a field. These same properties are used to define other systems like **groups** and **rings**. Students who will go on to take more mathematics courses are well advised to commit these properties to memory.[3]

It is important to realize that the two binary compositions can be different for different fields. This should not be surprising. We already know that adding two rationals involves more work than adding two integers. The binary composition $+$ is defined differently for the integers and the rationals. A list of some different fields is given below:

1. The rational number system $\mathbb{Q}$
2. The real number system $\mathbb{R}$
3. The **complex** number system $\mathbb{C}$. These are all numbers that can be written in the form $(x+yi)$ where $x, y \in \mathbb{R}$ and $i = \sqrt{-1}$. Here the composition $+$ is defined to be

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

[3]Mathematicians work very hard to reduce the amount of memorization to a bare minimum. Like most people, they **hate** to sit and learn things by rote. Therefore we can be assured that **everything** in a university-level course is important.

and the composition $*$ is defined as

$$(a + bi) * (c + di) = (ac - bd) + (bc + ad)i$$

Examples would be $(3 + 4i) + (1 + 2i) = (4 + 6i)$ and $(3 + 4i) * (1 + 2i) = (-5 + 10i)$.

4. The **binary** number system which is composed of the two numbers 0 and 1 and $+$ is defined as $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$ and $\mathbf{1} + \mathbf{1} = \mathbf{0}$. The composition $*$ is defined as $0 * 0 = 0 * 1 = 1 * 0 = 0$ and $1 * 1 = 1$.

All the rational numbers are part of the reals and all the reals are part of the complex numbers. The compositions for each of these fields can be viewed as extensions of the compositions for the fields they contain. Adding two complex numbers that are also real numbers using the complex $+$ composition will give the same result as using the real $+$ composition. This is not true for the field of binary numbers since $1 + 1 \neq 0$ in the rational, real and complex number systems.

A field that is composed of at least two of the scalars from another field and has exactly the same binary compositions is called a **subfield** of that field. An example of a subfield of the complex field would be all numbers of the form $(x + yi)$ where x and y are rational (not real) numbers. This subfield and the rational field are two of the fields which are represented in the Lampp program, which accompanies this text. Very interesting results and applications are in store for the patient and careful reader. We begin this exploration with a theorem.

A **theorem** is a statement which we **prove** to be true. Most of mathematics is involved with developing theorems and proving them. Students learn to do this by studying existing theorems and examining their proofs, just as artists learn by making copies of masterpieces or children become skilled at computer games by watching how more advanced players do things and trying to emulate them.

The theorem we will examine has to do with trying to know what conditions we need to ensure that we have the correct scalars from a field to form a subfield. (Do not expect to read through the proof in one go. Take your time and make sure you understand each line before proceeding to the next. Write out the proof on scrap paper without looking at the text. This theorem is not especially easy and, if you can master it, you should be able to master every other proof in the text. If you still have difficulties then try rereading this section. If that doesn't help just go to the next section and come back here later.)

Suppose we have a collection of scalars $\mathbb{S}$ which we believe to be a subfield of a field $\mathbb{F}$. We must have that all of the nine properties from the definition of a field hold. The first five properties hold for all scalars in $\mathbb{F}$ so they must hold for the scalars in $\mathbb{S}$. They are, after all, the same scalars, just fewer of them, and the binary compositions are exactly the same. We must make sure, though, that the compositions result in scalars that are also in $\mathbb{S}$. (We must make sure that $\mathbb{S}$ is closed under the two binary compositions.) The remaining properties must also be checked because they involve assertions that certain scalars must exist in $\mathbb{S}$. We can do all of this by using the following theorem.

Theorem 1.1. *Two or more scalars from a field $\mathbb{F}$ can form a subfield $\mathbb{S}$ if and only if whenever $x, y \neq 0$ are in $\mathbb{S}$, then so are $x + (-y)$ and xy^{-1}.*

Proof. We start with the situation where we know that if we have $x, y \neq 0 \in \mathbb{S}$ then we also have that $x + (-y) \in \mathbb{S}$ and $xy^{-1} \in \mathbb{S}$. In other words, we know that if we have two scalars, one of which is not zero, from $\mathbb{S}$ (and they could even be the same scalar) then the scalars we get from evaluating the two expressions are also in $\mathbb{S}$. We need to show that this implies that properties 6-9 are true.

We have that $x + (-y) \in \mathbb{S}$. There is nothing to prevent us from saying that x and y refer to the same scalar, as long as that scalar is not zero. So let's do that. Let us say that $x = y$. Then we know that $x + (-x) \in \mathbb{S}$. But $x + (-x) = 0$ so 0 must be in $\mathbb{S}$. Similarly, $x * x^{-1} = 1 \in \mathbb{S}$. Since 1 is in $\mathbb{S}$, we can then have that $1 * x^{-1} = x^{-1}$ is also in $\mathbb{S}$. Also, $0 + (-x) = -x$ is in $\mathbb{S}$. So we have shown that $0, 1, -x$ and x^{-1} belong to $\mathbb{S}$, which shows that properties 6-9 are true. As we have already seen, the other properties are inherited from $\mathbb{F}$ so we only need to show that $\mathbb{S}$ is closed under $+$ and $*$. We now know that if $x, y \in \mathbb{S}$ then we must have x and y^{-1} in $\mathbb{S}$. And this means that $x(y^{-1})^{-1} = xy$ is in $\mathbb{S}$. So if we have $x, y \neq 0 \in \mathbb{S}$ then we must have $xy \in \mathbb{S}$, which is what we mean when we say $\mathbb{S}$ is closed under multiplication. Similarly, $x, y \in \mathbb{S}$ implies x and $-y$ are in $\mathbb{S}$ which implies that $x - (-y) = x + y \in \mathbb{S}$ so $\mathbb{S}$ is closed under addition. This is enough to show that the expressions listed in the first five properties all produce scalars that are in $\mathbb{S}$. (You may have noticed that we only proved things for $x, y \neq 0$ and not about 0 itself, except for its existence. That's okay because we already know all about 0 because its properties are inherited from $\mathbb{F}$.) □

Theorems are usually of the form:

Given *Condition 1* then we will have *Condition 2*.

The *Condition 1* of our theorem was that whenever we had any two variables x, y such that y is not equal to zero and both represent scalars from $\mathbb{S}$, then it was always true that $x + (-y)$ and xy^{-1} are in $\mathbb{S}$. If this were true then we would also have *Condition 2*, that $\mathbb{S}$, made up of two or more scalars from $\mathbb{F}$, would form a subfield of $\mathbb{F}$.

The theorem says nothing about how we would get *Condition 1*. To use the theorem we would have to show that a collection of scalars has this condition. Then we would know that it was a subfield **without having to show that each property of a field held**.

When we try to prove a theorem it is often helpful to rewrite the theorem in this form. We also have to be very sure of what each term means. Writing out the definitions of each term often leads directly to a proof.

There also must be justification for each line in the proof. For example, when we said

> Then we know that $x + (-x) \in \mathbb{S}$.

we could do so because we knew this from *Condition 1*, which we assume to be true. All we did was change $-y$ to $-x$. We could do this because the condition did not say 'for any two **unique** scalars $x, y \neq 0$'.

Let's examine another theorem and its proof. This theorem allows us to extend the distributive law which we have referred to as the eighth condition (item 8 of Definition 1.3) for the scalars we need to form a field. What we would like to do is show that $c(x_1 + x_2 + x_3) = cx_1 + cx_2 + cx_3$ for scalars c, x_1, x_2, x_3 in a field. Actually, we would like to show this for *any* number of scalars. To do this, we will need a method of proof called **induction**.

Induction is easy to understand and usually straightforward to use. It is a very good technique to have in one's mathematical toolkit. It can be illustrated using the small blocks called dominoes. If we line up a row of dominoes along the floor so that each domino is close to another, then we can tip over the entire line of dominoes by pushing over the first one. Such a row of standing dominoes is shown, from the side, in Figure 1.1.

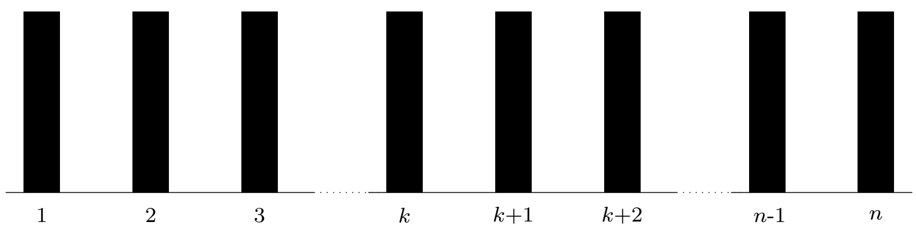


Figure 1.1: A Row of Dominoes

Notice that we do not show all the dominoes, just the first three, three somewhere in the middle and the last two.

Induction works on a list of n statements. Each statement is undeniably true or false. Also, each statement depends on its order in the list. One such statement might be that if we knock over the first domino, then the second will tip over. More generally, we might say that if domino k is knocked over, then it will knock over domino $k + 1$. Induction works in two steps.

Step 1. Show that statement 1 is true.

Step 2. Show that if statement k is true, then statement $k + 1$ is true.

If both of these steps are satisified, then we know that all n statements *must* be true.

If we show that if we knock over the first domino, the second domino will fall, we have shown Step 1. If we then show that *any* domino tipping over will cause the next in line to be knocked over, then we have satisfied Step 2 and we then know that all of the dominoes will be knocked over when we tip over the first one. Figure 1.2 illustrates these two steps.

Now we can use induction to prove our theorem.

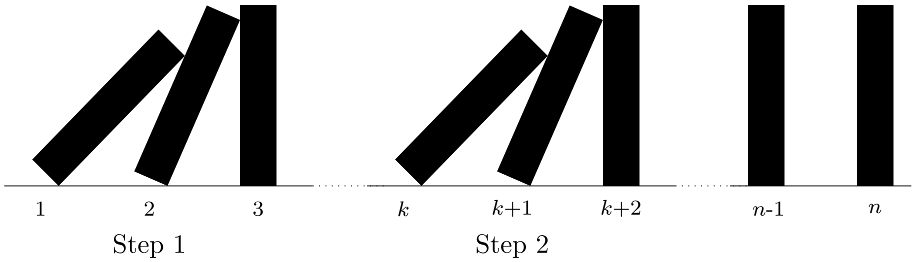


Figure 1.2: The Steps of Induction

Theorem 1.2. *Given a scalar c and n scalars x_i from a field, then*

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$$

Proof. We will accomplish the proof using the two steps of induction.

Step 1. This is easy. From the distributive law we have that $cx_1 + cx_2 = c(x_1 + x_2)$. Or

$$\sum_{i=1}^{2} cx_i = c\sum_{i=1}^{2} x_i$$

Step 2. We must show that if

$$\sum_{i=1}^{k} cx_i = c\sum_{i=1}^{k} x_i$$

is true, then

$$\sum_{i=1}^{k+1} cx_i = c\sum_{i=1}^{k+1} x_i$$

is true. We assume the first equation is true and add cx_{k+1} to both sides.

$$cx_{k+1} + \sum_{i=1}^{k} cx_i = cx_{k+1} + c\sum_{i=1}^{k} x_i$$

which is the same as

$$\sum_{i=1}^{k+1} cx_i = c\sum_{i=1}^{k+1} x_i$$

which is what we wanted to show.

By induction we have shown that

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$$

$\square$

Problems 1.3

1. Show that $\mathbb{Q}$ has no subfield other than itself. (We say that $\mathbb{Q}$ has no **proper** subfields, i.e. subfields that have less scalars than $\mathbb{Q}$.)

2. If you are really ambitious, show that the collection of scalars of the form $(x + yi)$ where $x, y \in \mathbb{Q}$ is a subfield of $\mathbb{C}$. Use Theorem 1.1.

3. Suppose $y \in \mathbb{C}$. Is $\sqrt{y} \in \mathbb{C}$?

4. Suppose we have two scalars a and b from a field. We can define scalar subtraction to be the binary composition of adding a to the additive inverse of b. In other words, $a - b = a + (-b)$. Since $1 \in \mathbb{F}$ and $-1 \in \mathbb{F}$ for any field $\mathbb{F}$, we can show that $(-1)(-1) = 1$ in any field.

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

   Show that $(-a)(-a) = a^2$ in any field $\mathbb{F}$.

5. Show that if we have a matrix C_n and a matrix $X_{n\times n}$ over $\mathbb{F}$, then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i x_{i,j} = \sum_{i=1}^{n} c_i \sum_{j=1}^{n} x_{i,j}$$

   (**HINT**: Use the theorem we proved that generalized the distributive law.)

Exercises 1.3

1. Evaluate $(2 + 5i) * (2 - 5i)$. (**HINT**: $i = \sqrt{-1}$ so $i * i = \sqrt{-1} * \sqrt{-1} = (\sqrt{-1})^2 = -1$)

2. Given the product $(3+xi)(1+4i) = y$ where $x \in \mathbb{R}$, what value must x have for y to be $\in \mathbb{R}$? What value must x have for y to be $\notin \mathbb{R}$? (We call any number $x = ai$ where $a \in \mathbb{R}$ a purely complex or **imaginary** number. The term imaginary is used for historical reasons and has absolutely nothing to do with the actual existence of such numbers. All numbers exist in the minds of people.)

1.4 Linear Equations

NOTE: Unless stated otherwise, all examples, exercises and problems in this text can be thought of as consisting of scalars from the rational or real fields.

In this section we will put together what we learned about matrices and fields with some concepts of algebraic geometry.

The equation

$$ax + by = c$$

is called a **linear equation** in two variables. The variables are x and y, one of which must not equal zero. The **constants** are a,b and c. They represent numbers (scalars) that do not change. Some examples are:

$$2x + 3y = 3$$

$$x + y = 2$$

$$y = -1$$

In the first example, $a = 2$, $b = 3$ and $c = 3$. In the second example, $a = 1$, $b = 1$ and $c = 2$. The third example has $a = 0$, $b = 1$ and $c = -1$. The variables x and y represent all scalars which make the equations true. In the second example, if $x = 1$ then $y = 1$. If $x = \frac{1}{2}$ then $y = \frac{3}{2}$. We can see that the solutions for these equations involve distinct pairs of scalars.

In geometry, all the scalars are real numbers. We write that $x, y, a, b, c \in \mathbb{R}$. Furthermore, geometrically, a linear equation represents a straight line in the plane. The variables x and y refer to the two **coordinates** of the plane. We write (x, y) to refer to a particular point on the plane. The variable x, the **x-coordinate**, refers to the horizontal position of a point. The variable y, the **y-coordinate**, refers to the vertical position. This is very similar to the way we refer to entries in matrices using row and column indices. The difference is that x and y refer to real numbers. Indices only use the counting numbers $1, 2, 3, \ldots$ or the integers ($\mathbb{Z}$).

All the pairs of scalars which are solutions for a linear equation are called the **graph** of the linear equation. We draw graphs of linear equations by putting in as many points as we need so that we might distinguish one equation from any other. Some examples of drawings of graphs are given below. (Even though it's not technically accurate, from now on we'll use the word *graph* to mean the drawing of a graph.)

The graph in Figure 1.3 on Page 18 represents a plane with the line formed by the linear equation $2x + 3y = 3$. The vertical line labeled **y** is called the ***y*-axis**. Similarly the horizontal line labeled ***x*** is called the ***x*-axis**. In order to draw the line we have to use a few of the algebraic properties of the real number field.

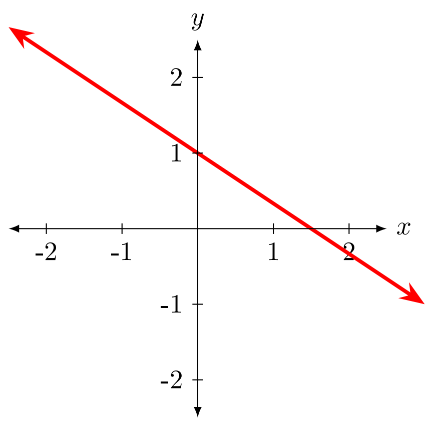


Figure 1.3: $2x + 3y = 3$

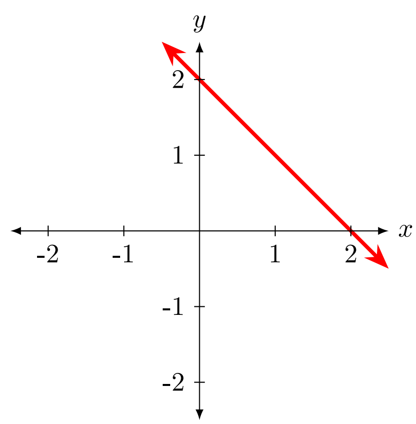


Figure 1.4: $x + y = 2$

We want the variable y to be alone on the left side of the equation. So first we subtract $2x$ from both sides.

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

This gives us

$$3y = -2x + 3$$

Then we multiply both sides by $\frac{1}{3}$.

$$\frac{1}{3} * 3y = \frac{1}{3} * (-2)x + \frac{1}{3} * 3$$

Notice that we had to use the distributive law on the right side of the equation.

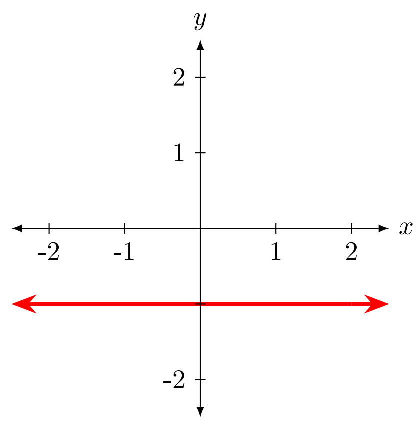


Figure 1.5: $y = -1$

This allows us to finally write

$$y = \frac{-2}{3}x + 1$$

Putting the equation in this form makes it easier to compute a y value for a given x (we call this **solving for *y***). For instance, let $x = 0$. Then $y = -\frac{2}{3} * 0 + 1 = 0 + 1 = 1$. We can see that the point $(0, 1)$ is on the graph of the equation. This point is called the ***y*-intercept** of the equation because it is where the line cuts the y-axis.

As you may have already surmised, there is also an ***x*-intercept**. We let $y = 0$ and solve either the original equation $2x + 3y = 3$ or the altered form $y = -\frac{2}{3}x + 1$. Let's stick $y = 0$ into the original equation.

$$2x + 3(0) = 3$$

$$2x = 3$$
$$x = \frac{3}{2}$$

So the point $(\frac{3}{2}, 0)$ is the x-intercept of the line.

In order to draw a graph we first draw the x and y axes (the plural of axis), mark the points and place a ruler lined up against them to draw the line. (Or we can use software to do it for us, which we will do with Lampp).

Instead of writing $2x + 3y = 3$ we could just write $\begin{pmatrix} 2 & 3 & 3 \end{pmatrix}$. In other words, we have a matrix with one row and three columns. All we need to keep in mind is that for a linear equation $ax + by = c$, the first column would refer to x and the second to y. The matrix $\begin{pmatrix} a & b & c \end{pmatrix}$ would be another way of representing a linear equation. (The constants a and b are often referred to as the **coefficients** of the linear equation and the constant c is called a **constant term**. The symbols x and y represent the variables.)

Of course we are not limited to two variables. A linear equation in n variables $x_1, x_2, \ldots, x_n$ is an equation that can be written in the form

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

where $a_1, a_2, \ldots, a_n$ and b are scalars of a field. We would write this in matrix form as $\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n & b \end{pmatrix}$. A solution of this equation would be a collection of scalars $s_1, s_2, \ldots, s_n$ that makes the equation true. That is:

$$a_1s_1 + a_2s_2 + \cdots + a_ns_n = b$$

We will see shortly that this method of representing linear equations is very useful. First, however, we will learn to use Lampp to graph lines and become familiar with this representation.

Problems 1.4

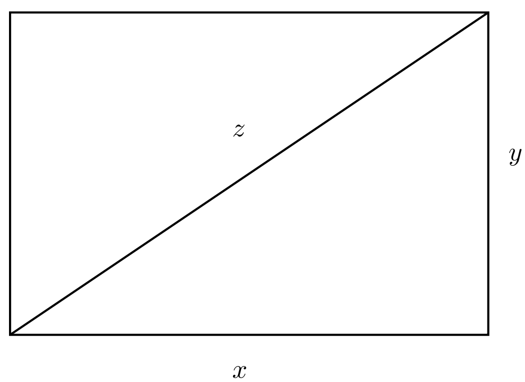


Figure 1.6: $x^2 + y^2 = z^2$

1. Suppose we have a linear equation $ax + by = c$ where all the coefficients, variables and the constant are scalars from the rational number system. We could also draw the graph of such an equation. However, not all numbers on the line would be represented. One such number was discovered by ancient Greeks when they studied the size of the diagonal of a **unit** square (a square with sides one unit long). They already knew that the length of the **diagonal** of a **rectangle** was related to the length of its sides using the famous theorem of **Pythagoras** which says that $x^2 + y^2 = z^2$ (see Figure 1.6). In other words, if $x = y = 1$ then $z^2 = 1 + 1 = 2$. (Since $z * z = 2$ we say that $z = \sqrt{2}$.) If the ratio of the diagonal to a side of a unit square was a rational number $\frac{m}{n}$ in its lowest form, we would have

$$\frac{m^2}{n^2} = 2$$

or that

$$m^2 = 2n^2$$

This would mean that m is **even** (divisible by 2). This means we could say that

$$m = 2p$$

and that

$$2n^2 = m^2 = 4p^2$$

and ultimately

$$n^2 = 2p^2$$

This means that n is divisible by 2. However, we assumed that $\frac{m}{n}$ was in its lowest form, so m and n can't have a common factor. This means our assumption about being able to have the ratio of the diagonal to the side as a rational number is impossible in this case. We say that $\sqrt{2}$ is **irrational** (not a rational number). Give examples of other irrational numbers.

2. In the preceding problem we assumed something and then showed that this assumption leads to a contradiction. Therefore our assumption is wrong and the opposite of what we assumed must be true. Mathematicians use this type of **proof by contradiction** all the time. Prove by contradiction that $x \neq \sqrt{x^2 + 1}$. (**HINT**: Assume that $x = \sqrt{x^2 + 1}$ and show that this leads to a contradiction.)

Exercises 1.4

1. Let $2x + 4y = 3$ be a linear equation in the reals. What is the x-intercept? What is the y-intercept?

2. If $ax + by + c = 0$ is a linear equation, what would be an equation for the x-intercept? The y-intercept? (**HINT**: In each case we know the value of one of the variables. Replace that variable with this number and solve for the remaining variable.)

3. Let $3x + 2y = 5$. Solve this for y. (**HINT**: We just want y by itself on the left of the equals sign.)

4. Let $ax + by + c = 0$. Solve this for y. The equation you develop is called the **point-intercept** form of the linear equation. What is the y-intercept of this line?

5. Let $3(x - 1) = \frac{1}{2}(y + 2)$. Is this a linear equation?

6. A man starts up a ramp that is 10 meters long and 2 meters high at the far end. Determine a linear equation that will give how high the man has risen for any horizontal distance he has traveled from the starting point. Use this equation to show how high he has moved when he is 8 meters from the start of the ramp. What is the actual distance he has traveled along the surface of the ramp? (**HINT**: Use the theorem of Pythagoras.)

7. A carpenter makes a wooden frame that is 3 meters by 4 meters. How can she use a tape measure and the theorem of Pythagoras to make certain that the frame is rectangular (all interior angles are 90°)?

1.5 Using Lampp to Graph Linear Equations

We have seen that we can represent a linear equation with a matrix of one row and three columns. We can graph a linear equation if it is composed of real or rational scalars. We shall do so now. We assume a working installation of the Lampp program on a PC or tablet. Information on obtaining the app can be found in the Appendix: *Lampp on your Computer.*

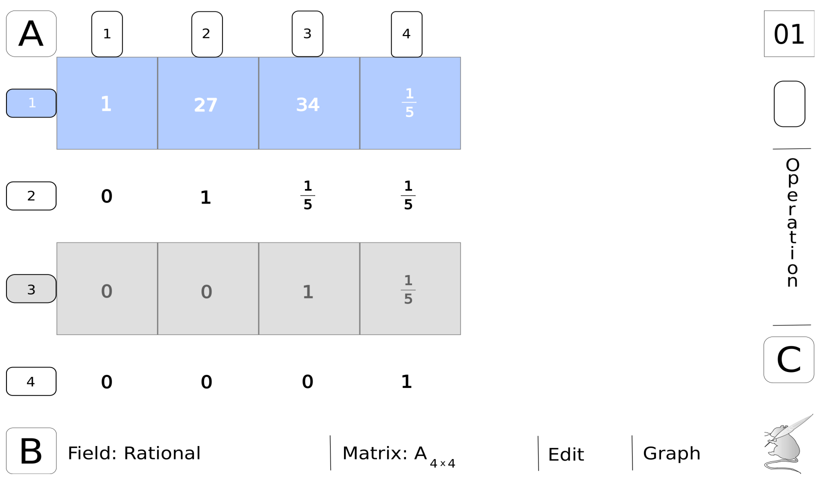


Figure 1.7: Lampp app with two rows selected

Depending on our device, we should see the Lampp app as something similar to Figure 1.7.

Activating features in Lampp is effected through selecting button or tab images by clicking mouse buttons or by touching or swiping on a touch screen. Clicking on the left mouse button when the cursor is at a certain location is usually identical to touching at that point. Right-clicking is equivalent to two touch screen actions: holding a touch and swiping right. The middle mouse button operation will usually accomplish the same effect as swiping downwards.

At the bottom middle of the Lampp screen is the Matrix tab. Left-click or touching this tab will cause the Matrix pane to slide up. You will see Figure 1.8. Use your device's keypad (real or virtual) to change the number of columns from 4 to 3. Activate the green Set control by left-clicking (or touching).

A confirmation dialog box may appear. If so, click OK. A matrix A of order 4 by 3 will be created (see Figure 1.9). Notice that some entries contain zero and some entries contain one.

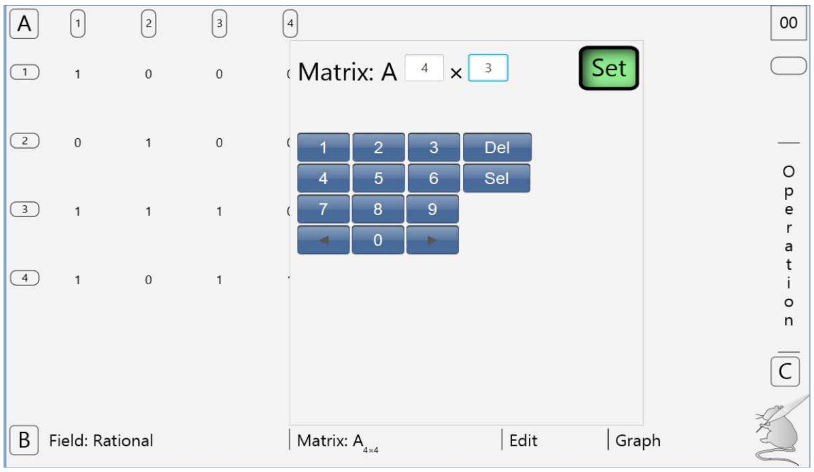


Figure 1.8: Creating a new matrix

Note that the default field is the rational numbers. The values of the entries correspond (in this case) to the leading coefficients of linear equations of the form: $a_i x + b_i y = c_i$, where i is the row number and the a,b,c's are, respectively, associated with the columns of the matrix.

Activating the Graph tab on the bottom right of the Lampp app will cause the Graph panel to slide up (see Figure 1.10). Each line is drawn for each row of the matrix. Clicking on the Line button on the bottom right of the Graph panel will cycle through each row equation.

In order to change the coefficients of the equations we are graphing, we need to use the Edit function. We select matrix entries to accept input by positioning the mouse cursor and left-clicking or simply touching the scalar we wish to change. The entry will become outlined and will now have a green background. Tapping or clicking again will deselect the entry.

After choosing one or more entries, select the Edit tab at the bottom of the screen. The scalar editor panel will slide up (see Figure 1.11). We can place numbers into the numerator or denominator of the Edit panel by first selecting for input with touch or the mouse or tab key and then typing in a value. Choosing the Set control will then have that value overwrite the selected matrix entries.

You should now feel free to experiment with graphing different linear equations. You should also check out the $*^{-1}$ and $+^{-1}$ buttons in the scalar editor panel to see what they do.

Exercises 1.5

1. Put a 1 in the first column of the matrix, a -1 in the second and a 0 in the third. Graph the linear equation.

2. Put a $\frac{1}{2}$ in the first column. Graph the equation. This scalar, the x

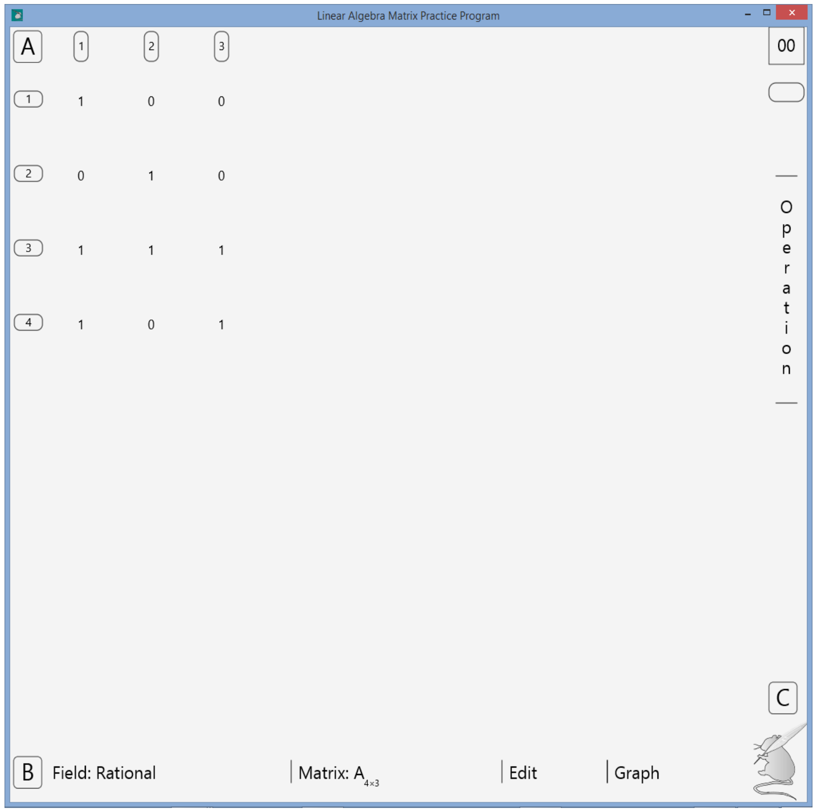


Figure 1.9: A new matrix

coefficient, is called the **slope** of the line. See what happens when you change its value. Write out a linear equation for this line.

3. Make $a_1 = 1$, $a_2 = 1$ and $a_3 = 0$. In other words, put a 1 in the first two entries and a 0 in the third entry. Graph the equation. Try changing the value of a_2 and see what happens.

4. Let $a_1 = 1$, $a_2 = -1$ and $a_3 = 0$. What happens to the line when the value of a_3 is changed?

1.6 Systems of Linear Equations

As remarked earlier, a linear equation in n variables is an equation of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$. When we have two or more of these equations, each one using the *same* variables $x_1, x_2, \ldots, x_n$, we have a **system** of m linear equations, where m is the number of such equations.

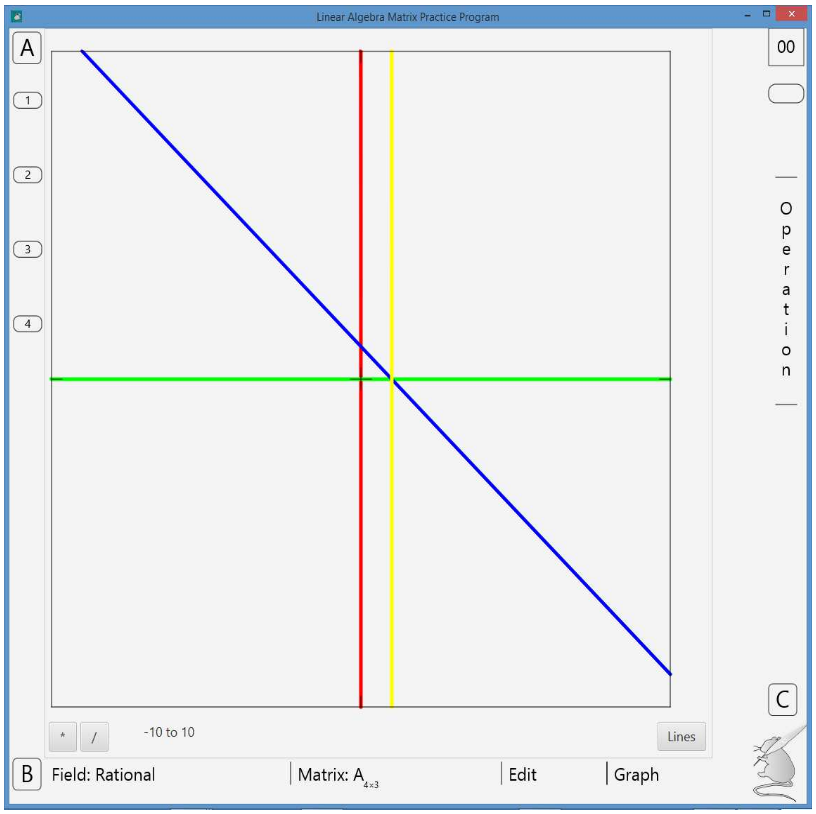


Figure 1.10: Calling up the Graph panel

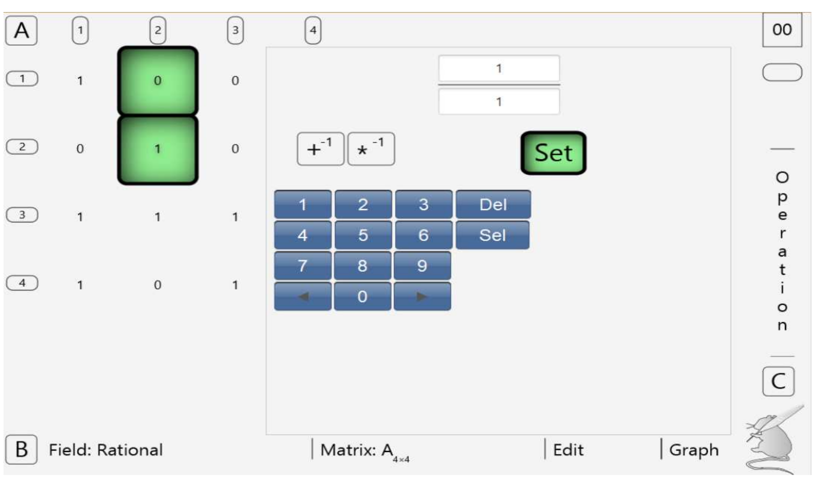


Figure 1.11: The Edit panel and two selected entries.

For the fields $\mathbb{R}$ and $\mathbb{Q}$ we can represent two-variable linear equations geometrically as lines on the plane. For example,

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

represents three lines on the plane. A solution of this system would be all the pairs of numbers (x, y) that satisfy all of the equations. Geometrically this would be all the points that the three lines have in common. This would happen when all three lines meet at a single point or all three equations represent the same line (the lines **coincide** or are **coincident**).

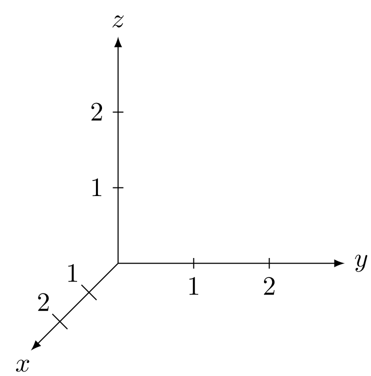


Figure 1.12: A projected 3-dimensional coordinate system

A solution for a system of equations can be obtained algebraically. For instance, suppose the first two equations in the example above were a system. We would take the first equation $x - y = 3$ and transform it into $x = y + 3$. Now we replace all occurrences of x in the second equation with $y + 3$. We would get $3(y + 3) + 2y = 0$ which we could write as $3y + 9 + 2y = 0$ or $y = -\frac{9}{5}$. We then take this value of y and put it in *either* of the two equations to get a value for x. The first equation would give us $x - (-\frac{9}{5}) = 3$ which we could manipulate to give $x = \frac{15}{5} - \frac{9}{5} = \frac{6}{5}$. (Using the second equation will give us the same answer.) So we know that the two lines **intersect** at the point $(\frac{6}{5}, -\frac{9}{5})$.

A linear equation in three variables x, y, z can be drawn as a **plane** in three-dimensional space. We add the z-coordinate to the x and y coordinates. This is easy to visualize. A box has width (x-axis), height (y-axis) and depth (z-axis). A two-dimensional representation of these three coordinates is given in Figure 1.12. A solution of a system of linear equations in three variables can be thought of geometrically as the point or points where all three planes meet. If they don't have points in common, we don't have a solution. A system of equations that has no solutions is called **inconsistent**. If it has one or more solutions then it is called **consistent**.

Geometry fails us when we deal with more than three variables. In fact it is not uncommon for people to deal with thousands of linear equations in thousands of variables. Linear algebra is used to discover if solutions to these systems exist, to find out if an existing solution is unique, and to discover how to calculate all the solutions.

Problems 1.6

1. Show that the system defined by

$$\begin{aligned} 5x + 10y &= 16 \\ x + 2y &= 3 \end{aligned}$$

   is inconsistent. What does this mean geometrically?

2. Geometrically, what are the two ways a system of linear equations with three variables can have solutions?

3. Give equations for three lines that intersect at one point. In other words, give a system of linear equations in two variables that have a unique solution.

Exercises 1.6

1. Use Lampp to create a 2×3 matrix corresponding to a system of the first two linear equations from the example in this section. Graph the two lines.

2. Create a 3×3 rational matrix with entries as in the first exercise using all three of the linear equations. Graph these three lines. Alter the values of the entries in the third row to try and get the three lines to meet at one point. What would be the linear equations for these lines?

3. Create a 1×4 matrix with the values $\begin{pmatrix} -1 & 1 & 1 & 0 \end{pmatrix}$. Graph this matrix. Make the first entry 1 and the second -1 and graph it again. Make the third entry -1 and the first two equal to 1. What does this graph look like?

4. Create a 2×4 matrix so that all the entries are non-zero. Graph the matrix. Try and alter the entries so that the two planes never meet.

5. Give an example of two linear equations for coincident lines. Use Lampp to graph these equations.

6. Solve the following system of equations:

$$\begin{aligned} 5x + 2y + 2z &= 14 \\ 4x - y + z &= 12 \\ 3x + z &= 9 \end{aligned}$$

(**HINT**: First find x in terms of z in the third equation. Then put this into the second equation. Then find y in terms of z in the transformed second equation. Then put the value of x and the value of y (both in terms of z) into the first equation. Solve this for z. Then use this number to calculate x and y).

7. Lampp interprets the first three entries of a row in a matrix with four columns as the variables for a 3-dimensional coordinate system. The linear equation $ax+by+cz = d$ is graphed as a portion of a plane in 3-dimensional space projected onto a 2-dimensional representation of a 3-dimensional box. Draw and label (with x, y and z) the coordinate system of the box.

   (**HINT**: Create a 1×4 matrix A with $a_{1,4} = 0$ and graph various planes.)

1.7 Linear Inequalities

When we defined the absolute value of a real number, we said that $|a| = a$ if a is greater than or equal to 0 and $|a| = -a$ if a is less than 0. Thus $|-7| = |7| = 7$, for instance. We will use the symbols $>$ and $<$ to stand for *greater than* and *less than*, respectively. Thus $4 > 2$ is true, but $-2 > 0$ is a false statement. We also use the symbol $\geq$ to mean *greater than or equal to* and the symbol $\leq$ for *less than or equal to*. Statements involving these four symbols are called **inequalities**. A **linear inequality** is a linear equation with one of the four inequality symbols used instead of an equals sign.

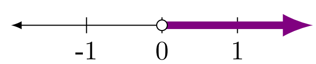


Figure 1.13: $a > 0$

A linear inequality in one variable can be described as a section of the line formed by the x-axis on the graph of a plane. For example, $a \geq 3$ would mean all the points to the right of 3 and the number 3 itself. The inequality $a > 3$ would look the same except that it would not contain 3. Figure 1.13 is what we usually would draw for a one-variable linear inequality. We use an unfilled circle, ○, at 0 on the line to indicate that a does not refer to zero. Figure 1.14 shows the inequality $a \leq 0$ where a can equal zero.

Figure 1.15 is an example of a linear inequality using the variable x as we would normally draw it on the x-axis. Figure 1.16 uses the variable y in the inequality $y \geq 1$ as it is drawn on the y-axis. We are free to draw on any line

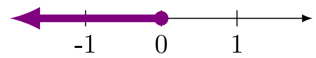


Figure 1.14: $a \leq 0$

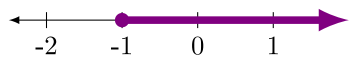


Figure 1.15: $x \geq -1$

in whatever direction we choose as long as we are careful to label our diagram properly. Usually, though, we stick to conventions and use the x-axis to draw one-variable inequalities.

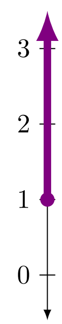


Figure 1.16: $y \geq 1$

As with linear equations, we can have a system of linear inequalities. In one variable this sytem gives a solution when the inequalities overlap. Figure 1.17 displays all the points that satisfies the system of inequalities given by $x > 2$ and $x \leq 5$. It is easy to see that 3 satisfies this system but that -1 or $7\frac{1}{2}$ would not.

Figure 1.17: $x > 2$ and $x \leq 5$

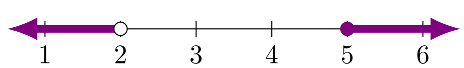


Figure 1.18: $x < 2$ or $x \geq 5$

When a system of inequalities does not have a solution, we say it is inconsistent. Figure 1.18 is an example of an inconsistent system of linear inequalities. We cannot have that x is less than 2 *and* be greater than or equal to 5 at the same time.

We sketch one-variable systems of inequalities on a line. Two-variable systems are drawn on a plane where each of the variables refers to an axis. The procedure to do this is quite simple. Suppose we had the linear inequality $y > x + 1$. First we would draw the line for the *equality* $y = x + 1$. Since y must be greater than this line, all the pairs of x and y points which satisfy the inequality would lie in the section of the plane above the line. A solution to a system of inequalities would be all the points in a region of a plane which satisfy all the inequalities, if such an area exists. We will examine one such system.

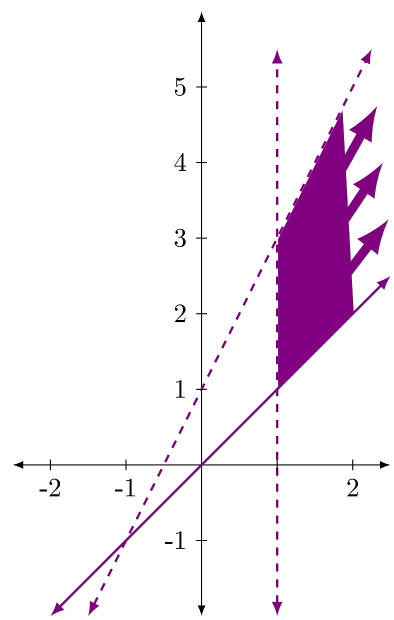


Figure 1.19: $y < 2x + 1$ and $y \geq x$ and $x > 1$

Figure 1.19 gives us the solution to the system of inequalities

$$\begin{aligned} &y < 2x + 1 \\ &y \geq x \\ &x > 1 \end{aligned}$$

We might start to draw such a graph by first drawing the line $x = 1$, which is parallel to the y-axis and cuts the x-axis at 1. Our inequality is $x > 1$ so we draw this line with dashes to indicate that points on this line are not to be included. Then we would shade the region of the plane to the right of this line. This shaded area contains all pairs of points (x, y) where x is greater than one.

We next draw the $y = x$ with a solid line because our inequality, $y \geq x$, contains all the points on this line. We pick a point, say $y = 1$ and $x = 2$, which falls below the line $y = x$. For these coordinates $1 < 2$, so obviously this point does not satisfy the inequality $y \geq x$. We then shade the region above the line $y = x$. Finally we draw a dashed line for $y = 2x + 1$ and shade the graph under this line (Why?). The area which is overlapped by all of this shading is the portion of the plane with points (x, y) which satisfy all of the inequalities.

Problems 1.7

1. We can rewrite the inequality $x + 1 > 2$ by adding -1 to both sides of the inequality sign. We would have $x + 1 + (-1) > 2 + (-1)$ which would give us the same results as $x > 1$. We can add or subtract the same term to both sides of an inequality without affecting it. The same can be said for multiplying each side of the inequality by a positive scalar. Show that multiplying both sides of inequalities by a zero or a negative scalar *does* affect the inequalities. How does it affect it?

2. Multiplying both sides of an inequality by a negative number switches the direction of the' inequality. Multiplying both sides of $y > 3$ by -1 would give us $-y < -3$. Multiplying both sides of an inequality by 0 will either give us a trivial or wrong result. Multiplying both sides by a positive scalar does not affect the inequality. Why, then, can we not multiply both sides of an inequality by a variable?

3. The absolute value of a variable in an inequality, $|x| < 1$, for instance, actually gives rise to a system of *two* inequalities, $x < 1$ and $x > -1$. What would be the two inequalities generated by $|x| \geq 3$? What would be the two inequalities generated by $|y + 3| > -3$? What about $|a| \geq |-3|$?

Exercises 1.7

1. Graph the following inequalities (**HINT**: Use Lampp to graph the *equalities*, redraw these on paper and shade the correct regions.)

   (a) $x + y \leq 2$ (b) $x \geq y - 2$

   (c) $2x_1 - x_2 > -3$ (d) $x < y + 3x + 1$

2. Graph the following systems of inequalities

   (a) $x < y + 5$
   $x \geq y$

   (b) $x > 2y - 1$
   $x - 4 \leq y$

(c) $y < x + 5$
$y > 3$
$x \geq -2$

(d) $x \geq y - 1$
$x > y$
$x < 5$
$y \geq 0$

3. Graph the following systems of inequalities

(a) $|x| > y - 3$

(b) $|y - 2| \geq 2y + x + 1$

(c) $y < \big|\,|x| - 2\,\big|$

(d) $-2x \geq -4y + 8$
$x < |y|$

Chapter 2

Row Operations

2.1 Gauss-Jordan Elimination

As we have seen, a linear equation can be represented as a matrix of one row. A system of equations can be written as a matrix with as many rows as there are equations. For example:

$$\begin{aligned} 5x + 2y + 2z &= 14 \\ 4x - y + z &= 12 \\ 3x \phantom{{}+2y} + z &= 9 \end{aligned}$$

could be written as

$$\begin{pmatrix} 5 & 2 & 2 & 14 \\ 4 & -1 & 1 & 12 \\ 3 & 0 & 1 & 9 \end{pmatrix}$$

Here the first column refers to the x coefficients, the second column to the y coefficients, the third to the z coefficients and the fourth to constant terms.

We would call a matrix made up of the first three columns a **coefficient matrix** and the matrix made up of just the last column a **matrix of constant terms**.

A general system of m equations in n variables,

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= b_2 \\ \vdots \qquad\qquad\qquad\qquad & \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= b_m \end{aligned}$$

is written as a $m \times (n+1)$ matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} & b_2 \\ \vdots & & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} & b_m \end{pmatrix}$$

By simply making sure that coefficients for each variable are entered only in the column assigned to that variable we save a lot of writing. This is not the only good reason to do this. It also makes it much easier to perform the algebraic operations needed to solve a system of equations. We do this by defining **row operations**.

Definition 2.1. ***Row Operations***

Operation ① *Interchange any two rows.*

Operation ② *Multiply all entries in a row by a nonzero scalar.*

Operation ③ *Add a multiple of one row to another row.*

Below, we give some examples of these three operations and show how they affect the corresponding system of linear equations.

$$\begin{pmatrix} 5 & 2 & 2 & 14 \\ 4 & -1 & 1 & 12 \\ 3 & 0 & 1 & 9 \end{pmatrix} \qquad \begin{aligned} 5x + 2y + 2z &= 14 \\ 4x - y + z &= 12 \\ 3x + z &= 9 \end{aligned}$$

We apply Operation ① to row 1 and row 2.

$$\begin{pmatrix} 4 & -1 & 1 & 12 \\ 5 & 2 & 2 & 14 \\ 3 & 0 & 1 & 9 \end{pmatrix} \qquad \begin{aligned} 4x - y + z &= 12 \\ 5x + 2y + 2z &= 14 \\ 3x + z &= 9 \end{aligned}$$

We demonstrate Operation ② by multiplying across the third row by $\frac{1}{3}$.

$$\begin{pmatrix} 4 & -1 & 1 & 12 \\ 5 & 2 & 2 & 14 \\ 1 & 0 & \frac{1}{3} & 3 \end{pmatrix} \qquad \begin{aligned} 4x - y + z &= 12 \\ 5x + 2y + 2z &= 14 \\ x + z/3 &= 3 \end{aligned}$$

We show Operation ③ by multiplying row 1 by -1 and adding it to row 2. *Note that we do* ***not*** *change row 1.*

$$\begin{pmatrix} 4 & -1 & 1 & 12 \\ 1 & 3 & 1 & 2 \\ 1 & 0 & \frac{1}{3} & 3 \end{pmatrix} \qquad \begin{aligned} 4x - y + z &= 12 \\ x + 3y + z &= 2 \\ x + z/3 &= 3 \end{aligned}$$

We will continue using row operations until we have the matrix of coefficients (the first three columns of our matrix) in the form:

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

One way to proceed would be to use Operation ② and multiply row 1 by $\frac{1}{4}$.

$$\begin{pmatrix} 1 & -\frac{1}{4} & \frac{1}{4} & 3 \\ 1 & 3 & 1 & 2 \\ 1 & 0 & \frac{1}{3} & 3 \end{pmatrix} \qquad \begin{aligned} x - y/4 + z/4 &= 3 \\ x + 3y + z &= 2 \\ x + z/3 &= 3 \end{aligned}$$

Now we will use Operation ③ again to add -1 times row 1 to row 3.

$$\begin{pmatrix} 1 & -\frac{1}{4} & \frac{1}{4} & 3 \\ 1 & 3 & 1 & 2 \\ 0 & \frac{1}{4} & \frac{1}{12} & 0 \end{pmatrix} \qquad \begin{aligned} x - y/4 + z/4 &= 3 \\ x + 3y + z &= 2 \\ y/4 + z/12 &= 0 \end{aligned}$$

And then we'll add -1 times row 1 to row 2.

$$\begin{pmatrix} 1 & -\frac{1}{4} & \frac{1}{4} & 3 \\ 0 & \frac{13}{4} & \frac{3}{4} & -1 \\ 0 & \frac{1}{4} & \frac{1}{12} & 0 \end{pmatrix} \qquad \begin{aligned} x - y/4 + z/4 &= 3 \\ 13y/4 + 3z/4 &= -1 \\ y/4 + z/12 &= 0 \end{aligned}$$

Let's add row 3 to row 1 (we are using Operation ③ and multiplying row 3 by 1 before we add it).

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} & 3 \\ 0 & \frac{13}{4} & \frac{3}{4} & -1 \\ 0 & \frac{1}{4} & \frac{1}{12} & 0 \end{pmatrix} \qquad \begin{aligned} x + z/3 &= 3 \\ 13y/4 + 3z/4 &= -1 \\ y/4 + z/12 &= 0 \end{aligned}$$

Then use Operation ② to get a 1 in column 2 of row 3 by multiplying by 4.

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} & 3 \\ 0 & \frac{13}{4} & \frac{3}{4} & -1 \\ 0 & 1 & \frac{1}{3} & 0 \end{pmatrix} \qquad \begin{aligned} x + z/3 &= 3 \\ 13y/4 + 3z/4 &= -1 \\ y + z/3 &= 0 \end{aligned}$$

Now we use Operation ③ and add -3 times row 3 to row 2.

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} & 3 \\ 0 & \frac{1}{4} & -\frac{1}{4} & -1 \\ 0 & 1 & \frac{1}{3} & 0 \end{pmatrix} \qquad \begin{aligned} x + z/3 &= 3 \\ y/4 - z/4 &= -1 \\ y + z/3 &= 0 \end{aligned}$$

Now multiply row 2 by 4 (Operation ②).

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} & 3 \\ 0 & 1 & -1 & -4 \\ 0 & 1 & \frac{1}{3} & 0 \end{pmatrix} \qquad \begin{aligned} x + z/3 &= 3 \\ y - z &= -4 \\ y + z/3 &= 0 \end{aligned}$$

Subtract row 2 from row 3 (i.e. multiply row 2 by -1 and add it to row 3).

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} & 3 \\ 0 & 1 & -1 & -4 \\ 0 & 0 & \frac{4}{3} & 4 \end{pmatrix} \qquad \begin{aligned} x + z/3 &= 3 \\ y - z &= -4 \\ 4z/3 &= 4 \end{aligned}$$

Multiply row 3 by $\frac{3}{4}$.

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} & 3 \\ 0 & 1 & -1 & -4 \\ 0 & 0 & 1 & 3 \end{pmatrix} \qquad \begin{aligned} x + z/3 &= 3 \\ y - z &= -4 \\ z &= 3 \end{aligned}$$

Add row 3 to row 2

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{pmatrix} \qquad \begin{aligned} x + z/3 &= 3 \\ y &= -1 \\ z &= 3 \end{aligned}$$

Finally, multiply row 3 by -$\frac{1}{3}$ and add it to row 1.

$$\begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{pmatrix} \qquad \begin{aligned} x &= 2 \\ y &= -1 \\ z &= 3 \end{aligned}$$

We can see that by using the row operations we have solved the system of equations. We check this by trying these values in our original equations

$$5 * (2) + 2 * (-1) + 2 * (3) = 10 - 2 + 6 = 14$$

$$4 * 2 - (-1) + (3) = 8 + 1 + 3 = 12$$

$$3 * (2) + (3) = 6 + 3 = 9$$

Which row operations we use to follow this procedure (called **Gauss-Jordan elimination**) and it what order is mostly a matter of taste. We usually try and pick the row operations to perform that will minimize the amount of calculation we need to do. The preceding example could certainly have been done with fewer row operations.

Problems 2.1

1. Do we really need three row operations? Operation ①, where we interchange two rows, could be replaced by applying the two other row operations. Show that this is true for any 2×2 matrix. Show that this is true for any $2 \times n$ matrix. (**HINT**: Remember that an entry of the ith row in a matrix A, for example, may be referred to as $a_{i,k}$ and that an entry in the same column in the jth row is $a_{j,k}$.)

2. Show that it is possible to replace Operation ② using only Operation ③ for any $3 \times n$ matrix where all entries in row 3 are zero.

3. Prove that Operation ① can be replaced by using the other two row operations for any size matrix as long as the matrix has more than 2 rows and it contains a zero row.

4. in Chapter 1 we talked about postfix notation. Invent a postfix notation to represent row operations.

2.2 Row Operations and Lampp

To the right of a matrix created with Lampp, an Operation panel can be opened. It is used to execute one of the operations on the rows selected with the right and left mouse buttons (or by touching or touch-and-holding, respectively).

In the matrix display, the matrix is indexed by the row of numbers in round-cornered box icons both above and to the left of the entries. By using the left mouse button to click on one of the left icon indices, that row will become blue to indicate that it has been selected. Touching will do the same on a touch-screen. We will call this **left-selecting** the row. The corresponding number of that row will be transferred to icons in the Operation panel. By clicking the right mouse button on a row icon (or touching-and-holding for more than 1 second), that row will become gray and the associated row icon in the Operation editor will be given that row's index. We will call this **right-selecting**. The radio buttons to the left of each operation listed in the Operation panel are used to select which operation to perform when the Set button is clicked. Figure 2.1 shows this panel.

The scalar entries in the Operation panel can be modified in the same way as the entries in the displayed matrix. Non-zero values in the scalar of the Edit panel are automatically sent to overwrite Operation scalars by clicking on the Edit panel Set button.

Just before an Operation is performed the value of the matrix is first saved and the value of the button at the top right is incremented. One may scroll through the values of the matrices for each operation performed on the matrix using this button. With a mouse, clicking on the top, left corner will show the result of the previous operation. Clicking on the bottom right corner will increment the matrix shown (if possible). Clicking on the top right will show the base setting of the matrix; the matrix before any operation was executed. Clicking on the bottom left will show the matrix after the last operation executed.

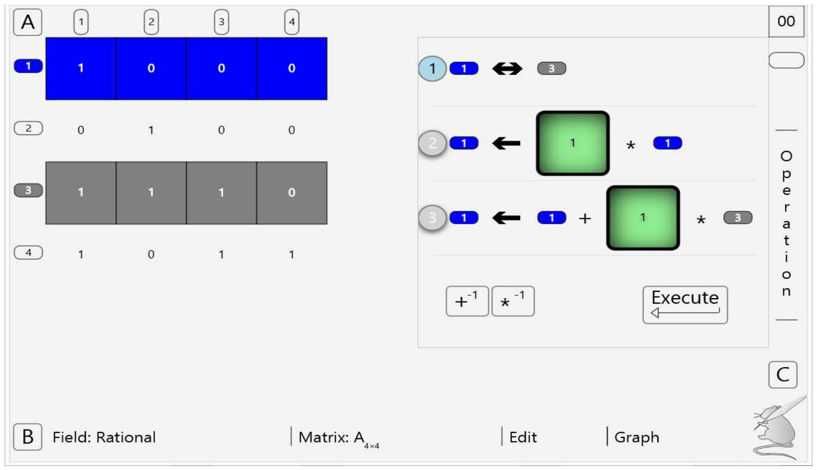


Figure 2.1: The Operation pane

On a touch screen, these actions are accomplished by swiping down on the button, swiping up, swiping left and swiping right, respectively.

Exercises 2.2

1. Use Lampp to follow the example of Gauss-Jordan elimination given in this chapter.

2. Use Lampp to find the solution of the example using less row operations.

3. The original equations in the example had only integer coefficients and integer constant terms. The solution values were also integers. Can you find a way to solve the system that doesn't involve any pure fractions (rationals that are not integers)?

4. Use Lampp to graph each step of the example. Then graph each step of the quicker method you found. What can you say about the difference between them?

2.3 Reduced Row Echelon Form

Definition 2.2. *If a matrix has the following two properties, it is in* ***row echelon form****.*

1. *Any row consisting entirely of* 0*'s appears at the bottom of the matrix.*

2. *For any two consecutive rows with one or more nonzero entries, the leftmost nonzero entry (also called the* ***leading entry*** *or* ***pivot****) of the lower row is to the right of the upper row leading entry.*

Definition 2.3. *A matrix already in row echelon form is in* ***reduced row echelon form*** *if*

1. *In any row, the leading entry is a* 1*.*
2. *Any column containing a pivot has zeros for entries everywhere else in the column.*

Once we have a few examples these definitions will not seem so daunting. The following two matrices are in reduced row echelon form.

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

Some examples of matrices that are **not** in reduced row echelon form are

$$\begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 9 \end{pmatrix}$$

We can see that the example on the left has a 3 in the column containing a leading entry, so this violates condition 2 of Definition 2.3. This matrix is, however, in row echelon form. The example on the right violates condition 2 of Definition 2.2 as the leading entry in row 2 is further to the right of the leading entry in row 3. This matrix is not even in row echelon form and therefore cannot be in reduced row echelon form.

Another phrase used for a matrix (or system of equations) in reduced row echelon form is **row canonical form**.

Since a field always has a 1 and a 0, we can see that the definition of reduced row echelon form will apply to any matrices in any field. Let's look at an example[1] of a matrix in the binary field.

Recall that the binary field consists solely of the numbers 0 and 1 and obeys the same laws of addition and multiplication as do the integers except that $1 + 1 = 0$. We will use these scalars to illustrate how use row operations to put a matrix in reduced row echelon form.

We will create $A_{4\times 5}$ with scalars $a_{i,j}$ from the binary field. The matrix will represent a grid of people. Each row will represent a group of egg-throwers. Each member of the group is able to hold at most one egg in their hands at a time. If someone throws them an egg, they will drop their own and fail to catch the incoming egg. The eggs are **raw**.[2] If somebody throws their egg, they will immediately be able to pull a new egg out of their pocket.[3] So, any member of a group will either have one egg in their hands or none. They can only throw an egg if they are holding one. They can also throw an egg if they have one

[1] A very artificial example.
[2] This is not important, just more fun to imagine.
[3] It's magic.

in their hands and be able to immediately get another. Otherwise, the only time anybody can gain an egg is if they aren't holding one already and someone throws them one.

Each group is made up of five characters: Larry, Moe, Curly, Shemp and Emil. We can refer to them by subscripts. For example, Larry_2 would be the Larry in row 2. We further add the condition that a character can only throw an egg to another character with the same name. Only a Moe can throw to a Moe, for instance.

We also stipulate that all members in a row will throw any egg they have the same distance at the same time. If Shemp_3 throws his egg to Shemp_1, then Larry_3, Moe_3, Curly_3 and Emi_3 will throw their eggs, if they have one at hand, to Larry_1, Moe_1, Curly_1 and Emil_1, respectively.

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \begin{array}{l} \text{A Larry, Shemp and Emil in this row each have an egg.} \\ \text{The Moe, Curly and Emil in this row each have an egg.} \\ \text{Only Moe in this row doesn't have an egg.} \\ \text{This row's Moe, Curly, Shemp and Emil have eggs.} \end{array}$$

Now let's use the $+$ and $*$ compositions for the binary field to put this matrix in reduced row echelon form. First we add row 1 to row 3 (we use Operation 3 to first multiply row 1 by 1 and then add this row to row 3). This means that everyone in row 1 will throw their eggs to their namesake in row 3.

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \begin{array}{l} \text{A Larry, Shemp and Emil in this row each have an egg.} \\ \text{The Moe, Curly and Emil in this row each have an egg.} \\ \text{Curly in this row keeps his egg.} \\ \text{This row's Moe, Curly, Shemp and Emil have eggs.} \end{array}$$

Then we add row 2 to row 4 (three eggs get tossed, six eggs get broken!)

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \begin{array}{l} \text{A Larry, Shemp and Emil in this row each have an egg.} \\ \text{The Moe, Curly and Emil in this row each have an egg.} \\ \text{Curly in this row keeps his egg.} \\ \text{Shemp in this row keeps his egg.} \end{array}$$

Our final two row operations would be to add row 3 to row 2 and to add row 4 to row 1.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \begin{array}{l} \text{A Larry and an Emil in this row each have an egg.} \\ \text{A Moe and an Emil in this row each have an egg.} \\ \text{Only Curly in this row has an egg.} \\ \text{Only Shemp in this row has an egg.} \end{array}$$

A little thought assures us that this final matrix gives us the minimum number of eggs that can be held at any point in the egg-hurling process, given the initial conditions.

This is a trivial (humorous?) example of the use of binary numbers to model a 'game'. But actual problems similar to this occur in the design of circuits and networks (although the problems are enormously more complicated and generally involve solution techniques beyond linear algebra.) It is possible for companies to save millions of dollars in production costs by identifying the minimum conditions necessary to obtain certain results.

In each step of the Gauss-Jordan elimination we come up with a new matrix. We say that a matrix is **row equivalent** to another matrix if it is possible to produce one matrix from another using a finite number of row operations.

Problems 2.3

1. Write out a procedure for using row operations to put a matrix in reduced row echelon form. You might start with:

   (a) Move all zero rows (rows with only 0 in its entries) to the bottom of the matrix using Operation 1.

2. For a matrix $R_{m\times n}$ in reduced row echelon form and where $m > n$, what is the least number of zero rows that R can have?

3. If we had the matrix
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$
   where $ad - bc \neq 0$, show that the reduced row echelon form is
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

4. If we had a matrix composed of only nonnegative real numbers, each of whose rows and columns sums to 1, why would such a matrix have to be square? [4]

5. **Column operations** are the same operations as row operations except they are performed on the columns of a matrix. When would a matrix in reduced row echelon form be equal to the matrix put in reduced column echelon form? In other words, if A is row equivalent to B, then can A be column equivalent to B?

6. Do you think the reduced row echelon form of a matrix is unique?

[4]In mathematics, especially in probability and combinatorics, such a matrix is called a **doubly stochastic matrix** (also called **bistochastic**). Can you think what just a **stochastic matrix** would be?

Exercises 2.3

1. Put the following matrices in reduced row echelon form. Assume that all entries of $A, B, C \in \mathbb{Q}$.

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 2 & 0 \\ 1 & 3 & 1 \\ 4 & 0 & 1 \\ 0 & 3 & 0 \end{pmatrix} \qquad C = \begin{pmatrix} 2 & 2 & 0 & 1 \\ 1 & 3 & 1 & 0 \end{pmatrix}$$

2. Use Gauss-Jordan elimination to solve the following systems of equations

$$\begin{aligned} 3x - 2y &= 25 \\ 2x - 3y &= 25 \end{aligned} \qquad \begin{aligned} 9x - 6y &= 75 \\ 2x - 3y &= 25 \end{aligned} \qquad \begin{aligned} 2x - 3y &= 25 \\ 9x - 6y &= 75 \end{aligned}$$

3. In regards to the matrices associated with the systems of linear equations in the previous problem, what row operations were used to transform the first matrix into the second? The second into the third?

4. Use Lampp to create the binary matrix of the example of the pie fight. To do this click on **Field** in the menu strip and select **Integer Modulus a Prime**. Accept the default for the prime modulus, which is 2. Create the matrix as you would for the rational field. Send values from the scalar Editor panel and use the $+^{-1}$ and $*^{-1}$ buttons to populate the matrix. Then use the Operation panel in order to put the matrix in reduced row echelon form. Can you reduce the matrix without using Operation 3? Why or why not?

5. The following problem comes from a Chinese text over 2000 years old. Use Gauss-Jordan elimination to solve it.

> There are three types of corn, of which three bundles of the first, two of the second, and one of the third make 39 measures. Two of the first, three of the second and one of the third make 34 measures. And one of the first, two of the second and three of the third make 26 measures. How many measures of corn are contained of one bundle of each type?

2.4 Parameters

We have seen previously that some systems of equations can have many solutions. For the rational, real and complex fields, we have either only one or an infinite number of solutions if any solutions exist. For example, when we had intersecting lines we had only one solution, the point where they crossed. If the lines were coincident then any point on either line was a solution. In this case either equation could be used to express the solutions for this system. Things got a little more complicated when we talked about planes intersecting. We saw that two planes could intersect at a point or a line. The planes could also be coincident. When we have an infinite number of solutions, like when two planes intersect at a line, we give an equation as an answer, in this case the equation of the line.

In order for a system of linear equations to have a unique solution we must have the same number of variables as we have equations. This is a necessary condition but it is not a sufficient condition. A system of linear equations with the same number of variables as equations could still have an infinite number of solutions. How do we know when a solution is unique? How do we express the solutions if there are an infinite number?

We start to answer the above questions by introducing the **rank** of a matrix.

Definition 2.4. *Suppose we have a matrix A and another matrix R in reduced row echelon form such that A is row equivalent to R, then the rank of A is simply the number of rows of R that contain leading entries.*[5]

If we had an $m \times (m+1)$ matrix D such that the rank of D equaled m, then we would have a unique solution given by column $(m+1)$ of the reduced row echelon matrix that was row equivalent to D.

If the rank is smaller than m, we have many solutions. Let's consider one such case. Say we have the system

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

This system represents two planes. Its associated matrix would be

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

This matrix takes the reduced row echelon form

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

Which corresponds to

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

We can use a little algebra to solve for x and y in terms of z.

$$x = \tfrac{1}{2}z + \tfrac{1}{2} \ \text{ and } \ y = -\tfrac{4}{3}z - \tfrac{1}{3}$$

[5] This is not a very good definition but it is accurate enough for our present purposes. A general definition, involving more advanced concepts, will be given later as Definition 5.6.

For any given z we can now easily calculate x and y values. Say we have an arbitrary scalar t (in this case $t \in \mathbb{R}$ since we are talking about planes). Let $z = t$ and we have the solutions

$$x = \tfrac{1}{2}t + \tfrac{1}{2}\ , \quad y = -\tfrac{4}{3}t - \tfrac{1}{3} \ \text{ and } \ z = t$$

This variable t is called a **parameter**.

Of course the solutions for some systems can have more than one parameter. Consider the system

$$\begin{aligned} x_1 \quad\; - 2x_3 + x_4 &= 5 \\ x_2 + 3x_3 + x_4 &= 8 \end{aligned}$$

We can allow x_3 and x_4 to be any numbers in the reals (which happens to be the field we are interested in at the moment). To give the solutions we must introduce two parameters s and t which represent any real numbers. We let $x_3 = s$ and $x_4 = t$ so that we get

$$\begin{aligned} x_1 &= 5 + 2s - t \\ x_2 &= 8 - 3s - t \end{aligned}$$

We can write out the solution as a matrix of one column

$$\begin{pmatrix} 5 + 2s - t \\ 8 - 3s - t \\ s \\ t \end{pmatrix}$$

where the row indices correspond to the indices of the variables x_i.

Problems 2.4

1. Given a matrix $A_{m \times n}$, give a formula for the number of parameters you would need for a solution. (**HINT**: Assume that the rank of A is equal to r.)

2. Consider the system of linear equations

$$\begin{aligned} x \ + \quad\; y &= 1 \\ a * x \ + \ a * y &= k \end{aligned}$$

   where $a > 1$ and k are constants. What value must k have for the system to have no solution? Infinitely many solutions? Can it have exactly one solution? (**HINT**: find the reduced row echelon form of the associated matrix. Whenever we have a reduced row echelon matrix equivalent to a linear system of equations such that there is a row which has all zeros for the coefficients but otherwise has one or more non-zero entries, we say the system is **inconsistent** and has no solution.)

3. Consider the system of linear equations

$$\begin{aligned} 2x + y &= 2a + b \\ -6x - 3y &= a - b \end{aligned}$$

where a and b are variables. What values must a and b have for the system to have solutions? (**HINT**: find the reduced row echelon form of the 2×4 associated matrix

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

Then interpret the non-zero entries of the bottom row as coefficients for an equation using a and b as variables. Show how a and b must be related for the system to be consistent.)

4. Assume that the following system is in the binary number field.

$$\begin{array}{rrl} a & & + c = 1 \\ & b & + c = 0 \end{array}$$

Give the list of parameters as a matrix. How many unique solutions are there? What are they?

Exercises 2.4

Solve the following problems by introducing parameters. Unless otherwise stated, assume that the systems are over the reals (all scalars $\in \mathbb{R}$). Give each solution as a matrix of one column.

1. $$\begin{array}{rrl} x & & +z = 4 \\ & y & +z = 2 \end{array}$$

2. $x + y + z = 1$

3. $$\begin{array}{rrl} 3x_1 & +2x_2 & +x_3 = 9 \\ & 3x_2 & +x_3 = 3 \end{array}$$

4. $$\begin{array}{rrrcr} 2x + & y + & 12z & = & 1 \\ x + & 2y + & 9z & = & -1 \end{array}$$

5. For each of the previous exercises use Lampp to graph the system. Choose values for the parameters in the solution and calculate the result. Notice how each result can be identified as the coordinates of a single point on the plane segments of the corresponding graphs. Try to identify the position of the point on the graph.

6. A wine company wants to create a product that has 12 percent alcohol by blending together three wines which have alcohol contents of 10, 16 and

13 percent, respectively. How much of each type of wine will go into one bottle? (**HINT**: give each of the volumes of the three wines a variable name, say a, b and c. So we know

$$\frac{10}{100}a + \frac{16}{100}b + \frac{13}{100}c = \frac{12}{100}$$

because this would give us the desired percentage of alcohol. We also know that

$$a + b + c = 1$$

because the total volume of the three wines must equal the volume of one bottle of the blended wine. Now use parameters to solve this system. Notice that the parameter you calculate has limits on its values in order to give sensible answers. What values can the parameter have?)

Chapter 3

Matrix Arithmetic

In some ways we have been rather fast and loose about what the associated matrix is for a system of linear equations. We will now introduce the properties of matrix arithmetic that will clear up the inconsistencies.

3.1 Matrix Addition and Scalar Multiplication

Definition 3.1. *For any two $m \times n$ matrices A and B, the **sum** of A and B is a matrix $C_{m\times n}$ with entries $c_{(i,j)} = a_{(i,j)} + b_{(i,j)}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$.*

One example should suffice. Suppose we had the two matrices

$$A = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 2 & 3 & 1 & 2 \\ 3 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 4 & 6 & 8 \\ 1 & 3 & 5 & 0 \\ 3 & 1 & 2 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}$$

Then $A + B = C$ would mean

$$C = \begin{pmatrix} 3 & 6 & 9 & 9 \\ 3 & 6 & 6 & 2 \\ 6 & 2 & 4 & 7 \\ 1 & 2 & 4 & 3 \end{pmatrix}$$

It is easy to see that matrix addition is commutative ($A + B = B + A = C$). It should be equally apparent that matrix addition is associative ($(A+B)+C = A+(B+C)$). If we had a matrix A and a second matrix of the same order with all of its entries equal to zero, the second matrix has all the properties of an identity for matrix addition. We reserve the bold symbol $\mathbf{0}$ for such a matrix and have $A + \mathbf{0} = \mathbf{0} + A = A$. We call $\mathbf{0}$ the **zero matrix**. The order of $\mathbf{0}$ is usually determined by its use in an equation to save us from having to write $\mathbf{0}_{m\times n}$.

We define the **product** of a **scalar** s and a **matrix** $A_{m\times n}$ as the matrix $B_{m\times n}$ such that $sA = B$ if and only if $s * a_{i,j} = b_{i,j}$ for any pair i and j such that $1 \leq i \leq m$ and $1 \leq j \leq n$.

If we have $(-1) * A = B$ then B is the unique additive inverse of A and $A + B = \mathbf{0}$.

Note that Operation 2 of our row operations is simply the product of a nonzero scalar and a matrix of one row.

Problems 3.1

1. How would you define matrix subtraction?

2. For a scalar r and two matrices A and B with orders such that addition is defined, is $r(A + B) = rA + rB$?

3. Show that $1A = A$ and that $0A = \mathbf{0}$.

4. Show that $r(sA) = (rs)A$.

5. If A and B have the same rank, is $A + B$ necessarily defined?

6. If $r \neq 0$, is the rank of A the same as the rank of rA?

7. If $A_{2\times 2}$ and $B_{2\times 2}$ are in row reduced echelon form, what can you say about the rank of $A + B$?

8. Show that if $A + B = C$ and $D + B = C$, then $A = D$.

Exercises 3.1

1. Assume that A and B are over $\mathbb{Q}$. Find $A + B$. Find $(-2)A$.

$$A = \begin{pmatrix} -2 & 1 & 4 \\ 2 & 3 & 5 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 4 & 4 \\ 1 & 3 & 2 \end{pmatrix}$$

2. Assume that A and B are over the binary field. Find $A + B$. Find $A - B$.

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

(**HINT**: You can use Lampp to do this by creating a 6×3 matrix with the first three rows being A and the bottom three rows being B. Then use row Operation 3 to add the rows of B to A.)

3. Assume that A and B are over $\mathbb{C}$. Find $A + B$. Find $3A$. Find iA. Find $(1 - i)B$. (Note that, in this case, $i = \sqrt{-1}$.)

$$A = \begin{pmatrix} 1 & i \\ 0 & 2 \end{pmatrix} \quad B = \begin{pmatrix} -1 + i & 4 \\ 1 - i & 1 - 2i \end{pmatrix}$$

3.2 Matrix Multiplication

Matrix multiplication is not so straightforward. We do not simply multiply each corresponding pairs of entries. We will use the $\sum$ notation we introduced earlier and give the complete definition. Then we will give special examples that will make the definition clearer.

Definition 3.2. *For two matrices $A_{m\times n}$ and $B_{n\times p}$ over some field $\mathbb{F}$, the **product** of A and B is a matrix $C_{m\times p}$ such that*

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} * b_{k,j}$$

for all pairs of i, j such that $1 \le i \le m$ and $1 \le j \le p$.

We would say that $AB = C$.

First notice that A has the same number of columns as B has rows. This means that BA is **not defined** unless $m = p$. Thus matrix multiplication is **not** commutative in general.

Let's look at an example when $m = p = 1$ and $n = 3$. Let

$$A_{1\times 3} = \begin{pmatrix} 2 & 3 & 5 \end{pmatrix}$$

and let

$$B_{3\times 1} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

then

$$AB = c_{1,1} = \sum_{k=1}^{3} a_{1,k} * b_{k,1} = (2*(-1)) + (3*1) + (5*0) = 1$$

Since we refer to any 1×1 matrix as just a scalar, we could write $AB = c = 1$. Multiplying any matrix of just one row and a matrix of just one column (when they are of the proper order for multiplication to be defined) will result in one scalar.

Let's redefine A and B so that

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

and

$$B = \begin{pmatrix} x \\ y \end{pmatrix}$$

We would then have

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{2} a_{1,k} * b_{k,1} \\ \sum_{k=1}^{2} a_{2,k} * b_{k,1} \end{pmatrix} = \begin{pmatrix} 1x + 2y \\ 3x + 4y \end{pmatrix} = C$$

It is apparent that c_1 is an entry made up of the matrix product of row 1 of A and column 1 of B (the only entry that column B has).

We can see now that the system of linear equations

$$\begin{aligned} x_1 - 2x_3 + x_4 &= 5 \\ x_2 + 3x_3 + x_4 &= 8 \end{aligned}$$

could be written as $AX = B$ where

$$A = \begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix} \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \text{ and } B = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$$

So that

$$\begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$$

A would be the matrix of coefficients and B would be the matrix of constant terms.

When we used Gauss-Jordan elimination, we put the matrix of coefficients together with the matrix of constant terms to create what is called an **augmented** matrix. Some people will draw a line between columns to distinguish the matrix of coefficients from the matrix of constant terms. This is a special example of a **partitioned** matrix, one where parts of the matrix are sectioned off to form **submatrices**.

Following are examples of an augmented and a partitioned matrix.

$$T = \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array}\right) \quad S = \left(\begin{array}{cc|cc} 2 & 4 & 6 & 8 \\ 1 & 3 & 1 & 4 \\ 3 & 9 & 2 & 2 \\ \hline 3 & 8 & 1 & 0 \\ 1 & 0 & 7 & 6 \end{array}\right)$$

The partitioned matrix S could also be written as a matrix with each submatrix indicated. For example

$$S = \begin{pmatrix} S_{1,1} & S_{1,2} \\ S_{2,1} & S_{2,2} \end{pmatrix}$$

Note that $S_{i,j}$ refers to a submatrix of S and $s_{i,j}$ refers to a scalar entry of S.

The use of augmented and partitioned matrices is a convenience we will take advantage of from time to time.

Matrix multiplication, like matrix addition, also has an identity. We give this **identity matrix** the special symbol $\mathbf{I}$. It has the properties that $A\mathbf{I} = A$ and $\mathbf{I}A = A$. Like the $\mathbf{0}$ matrix, $\mathbf{I}$ takes on the order necessary to make the preceding properties true. This was why we were careful to give both examples. For instance, if $A_{m\times n}$, then $A\mathbf{I} = A$ implies that $\mathbf{I}$ is, in this case, $n \times n$. If $\mathbf{I}A = A$, then $\mathbf{I}$ is $m \times m$. So what does $\mathbf{I}$ look like?

Definition 3.3. **I** *is an* $m \times m$ *matrix such that* $i_{j,j} = 1$ *for* $1 \leq j \leq m$ *and* $i_{j,k} = 0$ *for all* $j \neq k$ *and* $1 \leq j \leq m$ *and* $1 \leq k \leq m$*.*

An example would be

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We can see that **I** is a **square** matrix (a matrix with the same number of rows as it has columns) which has ones for its diagonal entries and is zero everywhere else.

For *some* square matrices A, there exists a matrix B such that $AB = BA = \mathbf{I}$. We call B the **inverse** of A. The inverse of a matrix is very important and deserves its own section.

Accomplishing matrix multiplication by hand can be very messy unless care is taken to eliminate arithmetic mistakes. We can define a procedure to multiply two matrices that involves a bit of writing but helps us keep track of each step. Suppose we wanted the product

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f & g \\ h & i & j \end{pmatrix}$$

Since we are multiplying a 2×2 matrix with a 2×3 matrix, the product will be a 2×3 matrix, a matrix with 6 entries. Let's create a table with 3 rows and 6 columns

$$\begin{array}{ccc|ccc} e & f & g & h & i & j \\ \hline a & a & a & b & b & b \\ c & c & c & d & d & d \end{array}$$

Multiply every entry in each column from the second row down by the top entry in its column

$$\begin{array}{ccc|ccc} e & f & g & h & i & j \\ \hline a*e & a*f & a*g & b*h & b*i & b*j \\ c*e & c*f & c*g & d*h & d*i & d*j \end{array}$$

Then for each row (other than the first), add the entry that is three columns over to each of the first three column entries. This is simply an example of matrix addition.

$$\begin{array}{ccc|ccc} e & f & g & h & i & j \\ \hline a*e+b*h & a*f+b*i & a*g+b*j & b*h & b*i & b*j \\ c*e+d*h & c*f+d*i & c*g+d*j & d*h & d*i & d*j \end{array}$$

The matrix made up of the first three columns of the bottom two rows is the product of the two matrices.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f & g \\ h & i & j \end{pmatrix} = \begin{pmatrix} a*e+b*h & a*f+b*i & a*g+b*j \\ c*e+d*h & c*f+d*i & c*g+d*j \end{pmatrix}$$

A numerical example would be

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

Write the rows of the right matrix across the top. Underline it. Then place the first column of the left matrix under each entry for the first row of the right matrix. Then place the second column of the left matrix under each of the entries for the second row of the right matrix. Then the same for the third column.

$$\begin{array}{cc|cc|cc} 4 & 6 & 7 & 5 & 1 & 4 \\ \hline 1 & 1 & 2 & 2 & 4 & 4 \\ 2 & 2 & 3 & 3 & 5 & 5 \end{array}$$

Multiply down each column by the topmost entry.

$$\begin{array}{cc|cc|cc} 4 & 6 & 7 & 5 & 1 & 4 \\ \hline 4 & 6 & 14 & 10 & 4 & 16 \\ 8 & 12 & 21 & 15 & 5 & 20 \end{array}$$

Then add the entries from column $i+2$ and $i+4$ to column i for $1 \leq i \leq 2$ for each of the rows other than the top row. Example: $4+14+4=22$.

$$\begin{array}{cc|cc|cc} 4 & 6 & 7 & 5 & 1 & 4 \\ \hline 22 & 32 & 14 & 10 & 4 & 16 \\ 34 & 47 & 21 & 15 & 5 & 20 \end{array}$$

This gives us the product matrix

$$\begin{pmatrix} 1 & 2 & 4 \\ 2 & 3 & 5 \end{pmatrix} \begin{pmatrix} 4 & 6 \\ 7 & 5 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 22 & 32 \\ 34 & 47 \end{pmatrix}$$

We can use Lampp to emulate this process. For instance, if we wanted the product

$$A_{n\times m}B_{m\times p} = C_{n\times p}$$

we would use Lampp to create a $(n+1) \times (m*p)$ matrix, putting in all the rows of B into the first row. The other n rows would be filled with p repeated columns of each column of A as in our examples. Then we would use **column operations** to accomplish the remaining calculations. Column operations are the same as row operations except that, of course, they act on columns. To get Lampp to use column operations, use the mouse (or your finger) to depress the row button just above the Operation tab.. The button will change to a column button. Use Operation 2 and Operation 3 as appropriate. Note that Operation 2 does *not* let you multiply by a zero scalar. You must accomplish this kind of multiplication by selecting all entries of a column which has a zero in the first row entry and setting them to zero. You can do this very easily by sending the zero to the Editor with the right mouse button (or right swiping), selecting the entries with the left mouse button (or touching) and then clicking the Set button in the scalar Editor.

Lampp is limited to creating matrices no bigger than 10 $\times$ 10 so it can only be used for multiplying small matrices. However, if you were stuck on a desert island without enough bamboo and coconuts to build a computer and if you had a burning need to multiply two matrices, you should now be able to do so.

Problems 3.2

1. Matrix multiplication is not commutative, but it is associative. In other words, $(AB)C = A(BC)$ provided all products are defined. We can show this by letting $A_{m\times n}B_{n\times p} = D_{m\times p}$ and $B_{n\times p}C_{p\times r} = E_{n\times r}$. We want to show that all of the entries of DC are identical to the entries of AE. We start by showing what the entries of D and E are.

   $$d_{i,j} = \sum_{k=1}^{n} a_{i,k} * b_{k,j} \qquad \text{Why? What values can } i \text{ and } j \text{ have?}$$

   Similarly,

   $$e_{i,j} = \sum_{k=1}^{r} b_{i,k} * c_{k,j}$$

   What values can i and j have in this statement?
   Then since $(AB)C = DC$ we can let $F = DC$ and write

   $$f_{i,j} = \sum_{s=1}^{p} d_{i,s} * c_{s,j} = \sum_{s=1}^{p}\left(\sum_{k=1}^{n} a_{i,k} * b_{k,j}\right) * c_{s,j} = \sum_{s=1}^{p}\sum_{k=1}^{n} a_{i,k} * b_{k,j} * c_{s,j}$$

   Why can we make this last statement?
   Similarly if we let $A_{m\times n}E_{n\times r} = G_{m\times r}$ we have

   $$\begin{aligned} g_{i,j} &= \sum_{k=1}^{n} a_{i,k} * e_{k,j} = \sum_{k=1}^{n} a_{i,k} * \left(\sum_{s=1}^{p} b_{i,s} * c_{s,j}\right) \\ &= \sum_{k=1}^{n}\sum_{s=1}^{p} a_{i,k} * b_{i,s} * c_{s,j} = \sum_{s=1}^{p}\sum_{k=1}^{n} a_{i,k} * b_{k,j} * c_{s,j} \end{aligned}$$

   Why can we interchange the $\sum$ signs in this last statement?
   We have shown that $f_{i,j} = g_{i,j}$ so that $(AB)C = A(BC)$.

2. If AB is defined, show that $A(rB) = r(AB)$.

3. Show that $A(B + C) = AB + AC$. Assume that all matrix sums and products are defined.

4. Let

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

   Show that if $AX = XA$ and $BX = XB$ for two matrices A and B, then $AB = BA$.

5. Show that it is impossible to have a matrix Y such that

$$XY = YX = \mathbf{I} \text{ when } X = \begin{pmatrix} -2 & -6 \\ 3 & 9 \end{pmatrix}$$

6. Evaluate the following matrix product

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

7. For the field $\mathbb{C}$, evaluate the following product

$$\begin{pmatrix} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i & 0 \\ 0 & \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i & 0 \\ 0 & \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i \end{pmatrix}$$

8. Given $A_{n\times n}$ and $B_{n\times m}$ with $m > n$.
Why can we always write $\left(A \,\middle|\, \mathbf{0}\right) = B$ but $\left(A \,\middle|\, \mathbf{I}\right) = B$ is only possible if $m = 2n$?

Exercises 3.2

1. Show how the following systems of equations can be written as arithmetical statements using matrices.

(a) $\begin{array}{rrr} x & & +z = 4 \\ & y & +z = 2 \end{array}$

(b) $x + y + z = 1$

(c) $\begin{array}{rrr} 3x_1 & +2x_2 & +x_3 = 9 \\ & 3x_2 & +x_3 = 3 \end{array}$

(d) $\begin{array}{rrr} x & & +z = 4 \\ x & +y & = 2 \end{array}$

2. The following matrices are over the field $\mathbb{Q}$. Find the products.

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

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

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

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

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

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

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

(h) $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

(i) $\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ (j) $\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

(k) $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 3 \end{pmatrix}$ (l) $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ 2 & 3 & 5 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

(m) $$\begin{pmatrix} 0 & 1 & 2 \\ -1 & 3 & 0 \\ 1 & -2 & 1 \end{pmatrix}\begin{pmatrix} -3 & 5 & 6 \\ -1 & 2 & 2 \\ 1 & -1 & -1 \end{pmatrix}$$

(n) $$\begin{pmatrix} -3 & 5 & 6 \\ -1 & 2 & 2 \\ 1 & -1 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 2 \\ -1 & 3 & 0 \\ 1 & -2 & 1 \end{pmatrix}$$

3. The following matrices are over the field $\mathbb{C}$. Find the products.

(a) $$\begin{pmatrix} 1+i & 1 & -1+i \end{pmatrix}\begin{pmatrix} 2-i & 3 & 1-2i \\ -1 & 1-i & i \\ -2i & 0 & 1 \end{pmatrix}$$

(b) $$\begin{pmatrix} i & 1-i \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2+i & 2i \\ 2 & 1-i \end{pmatrix}$$

4. The following matrices are over the binary field. Find the products.

(a) $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$

5. For a square matrix A, $A^2 = AA$ and $A^3 = AAA$. In general, A^n is called a **power** of A and is simply equal to the matrix A multiplied by itself n times. For matrices over $\mathbb{Q}$, find A^2 and A^3.

(a) $A = \begin{pmatrix} -6 & 12 \\ -3 & 6 \end{pmatrix}$ (b) $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -3 & -1 & -2 \end{pmatrix}$

6. For matrices over the binary field, find B^2 and B^3. What would B^n be for some positive $n \in \mathbb{Z}$?

   (a) $$B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

   (b) $$B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

7. For matrices over $\mathbb{C}$, find A^2 and A^3.

   (a) $$A = \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix}$$

   (b) $$A = \begin{pmatrix} 1 & 1+i & -i \\ -1+i & -1 & 1+i \\ i & -1+i & i \end{pmatrix}$$

8. For matrices X and Y over $\mathbb{Q}$, graph Y and XY.

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

   (a) $$X = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

   (b) $$X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

   (c) $$X = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

   (d) $$X = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{pmatrix}$$

9. For matrices X and Y over $\mathbb{Q}$, graph Y and XY.

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

   (a) $$X = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   (b) $$X = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   (c) $$X = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   (d) $$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

3.3 Matrix Inverses, Part A

It is common to think that jumping out of a hole is the inverse of jumping *into* a hole. The problem of performing this inverse action of getting out of a hole is usually more difficult than falling into it. Like most things in life, the more difficult action is the one we need to perform the most (or so it may appear). The same relationship holds for finding the inverse of a matrix.

Before arriving at the method for calculating the inverse of a matrix, let us look at one of the reasons we might find a matrix inverse useful.

In the previous chapter, row operations were used to solve systems of linear equations. For example, the system

$$\begin{aligned} 10x + 10y + 15z &= 45 \\ 20x - 5y + 5z &= 60 \\ 15x + 5z &= 45 \end{aligned}$$

could be written as

$$\begin{pmatrix} 10 & 10 & 15 & 45 \\ 20 & -5 & 5 & 60 \\ 15 & 0 & 5 & 45 \end{pmatrix}$$

When Gauss-Jordan elimination is used, we end up with

$$\begin{pmatrix} 1 & 0 & 0 & \frac{12}{5} \\ 0 & 1 & 0 & \frac{-3}{5} \\ 0 & 0 & 1 & \frac{9}{5} \end{pmatrix}$$

This gives us the solutions $x = 12/5$, $y = -3/5$ and $z = 9/5$.

If we let

$$A = \begin{pmatrix} 10 & 10 & 15 \\ 20 & 5 & 5 \\ 15 & 0 & 5 \end{pmatrix} X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} B = \begin{pmatrix} 45 \\ 60 \\ 45 \end{pmatrix}$$

then we can write our system of equations as

$$AX = B$$

In the section on matrix multiplication, we remarked that some square matrices (matrices with the same number of rows as columns) have inverses. We use the superscript $^{-1}$ to show that a matrix is an inverse. We say that A^{-1} is the inverse of A and that $A^{-1}A = AA^{-1} = \mathbf{I}$. We will use this as our definition.

Definition 3.4. *A square matrix A has an inverse A^{-1} if and only if $A^{-1}A = AA^{-1} = \mathbf{I}$.*

We now assume that A^{-1} exists for our system of equations and write

$$AX = B$$

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

$$\mathbf{I}X = A^{-1}B$$

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

Since X is a column matrix containing the variable identifiers x, y, z, we should easily be able to calculate values for these variables for different values for the scalars of B. That is, if we have A^{-1}, we just use matrix multiplication to find solutions for any specified constant terms of our system of equations *if solutions exist*.

If a square matrix A has an inverse, then A is called **invertible** or **nonsingular**. If it doesn't have an inverse, it is called **noninvertible** or **singular**. A matrix that is invertible is somewhat like a non-zero scalar.

In the next section we will introduce the concept of elementary matrices which will be used to develop a method to determine if an inverse exists and how to calculate it.

Problems 3.3

1. What is the inverse of $\mathbf{I}$?

2. Show that $\mathbf{0}$ has no inverse.

3. Assume that B and C are inverses of A. Prove that $B = C$. (This proves that the inverse of a matrix, if it exists, is **unique**.)

4. If A,B and the product matrix AB are all invertible, prove that

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

5. Suppose we have n invertible matrices A_1,A_2,. . .,A_n. Use the result of the previous problem to show that

$$(A_1A_2A_3 \cdots A_n)^{-1} = A_n^{-1}A_{n-1}^{-1} \cdots A_1^{-1}$$

Exercises 3.3

1. For the system of equations in the example of this section, show that

$$A^{-1} = \begin{pmatrix} \frac{-1}{45} & \frac{2}{45} & \frac{1}{45} \\ \frac{1}{45} & \frac{7}{45} & \frac{-2}{9} \\ \frac{1}{15} & \frac{-2}{15} & \frac{2}{15} \end{pmatrix}$$

2. For the system of equations in the example of this section, show that $X = A^{-1}B$.

3. What would be the values of x, y, z if

$$B = \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix}, \text{ if } B = \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix}, \text{ and if } B = \begin{pmatrix} 3 \\ 4 \\ 1 \end{pmatrix}$$

3.4 Elementary Matrices

Row operations, as we have seen, are extremely useful (and *fun*). They are also easy to use, especially when we have software like Lampp to take away some of the pain of arithmetic. However, it is difficult to *prove* much about the results of row operations in the form in which we actually use them. Luckily there is an equivalent way to represent row operations on a matrix using identity matrices, scalar multiplication and matrix multiplication. Since we have already started to prove properties of matrices, this representation will allow us to use our skills to prove results involving row operations.

A matrix obtained from an identity matrix by a single row operation is called an **elementary matrix**. We usually represent such a matrix as E. Since there are three row operations, we have three types of elementary matrices.

Definition 3.5. ***Elementary Matrices for Row Operations***

E Type ① *Interchange any two rows of* $\mathbf{I}$.

E Type ② *Multiply any row of* $\mathbf{I}$ *by a nonzero scalar.*

E Type ③ *Add a scalar multiple of any row of* $\mathbf{I}$ *to a different row.*

Examples of type ① elementary matrices:

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

Examples of type ② elementary matrices:

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

Examples of type ③ elementary matrices:

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

Multiplying a matrix *on the left* by an elementary matrix has the same effect as the equivalent row operation. In mathematical notation $EA = B$ means that B is the matrix A after one row operation. It is obvious that B is row equivalent to A.

Problems 3.4

1. Which of the following is an elementary matrix? If not, why not?

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

2. We had previously stated that $\mathbf{I}A = A$ for $A_{m\times n}$. Of course, $\mathbf{I}$ is $m \times m$ in this case. We can prove this using brute calculation. Since $\mathbf{I}$ is $i_{j,j} = 1$ and $i_{j,k} = 0$ whenever $j \neq k$, then if $\mathbf{I}A = B$, by the definition of matrix multiplication

$$b_{j,k} = \sum_{l=1}^{m} i_{j,l} * a_{l,k}$$

   Since $i_{j,l} = 1$ only if $j = l$ and otherwise it is zero, then $b_{j,k} = a_{j,k}$. So $B = A$ and $\mathbf{I}A = A$.

   Use a similar argument to show that $A\mathbf{I} = A$.

3. Show that **premultiplying** a matrix A by an elementary matrix E of type (1) corresponds to a type (1) row operation. (Premultiplying means to multiply on the left, i.e. EA.) Do the same for the other two row operations.

Exercises 3.4

1. Show that

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

   What are the corresponding row operations for each of these five elementary matrices?

2. If

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

   What is A^2? What is A^n if n is even? If n is odd? What does this mean in terms of row operations?

3.5 Matrix Inverses, Part B

If we interchange any two rows of $\mathbf{I}$ to get E and then interchange the same rows of E, we will obviously obtain $\mathbf{I}$ again. Since E is an elementary matrix of type (1), this means that $EE = \mathbf{I}$. Obviously E of type (1) is its own inverse.

An elementary matrix E of type (2) is identical to $\mathbf{I}$ except that $e_{i,i} = a$ for exactly one index value i and for some scalar $a \neq 0$. Let us say that the inverse of a type (2) E is equal to F. This means that $f_{i,i} = (e_{i,i})^{-1}$ for all values of i. It is easy to show that $EF = FE = \mathbf{I}$ so that $F = E^{-1}$. This also makes it easy to calculate E^{-1} for a type (2) elementary matrix. An example of a type (2) elementary matrix and its inverse are given below.

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad E^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{a} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

An elementary matrix of type ③ E is identical to $\mathbf{I}$ except that $e_{i,j} = a$ for some scalar a and for one specific index value i and one specific index value j and $i \neq j$. We say that E has one non-zero, **off-diagonal** entry. The **diagonal** entries, $e_{k,k}$, are all equal to one. All other entries are zero.

If we add a multiple of a row to another row and then subtract that multiple of a row from the same row, the matrix is left unchanged. This simple fact allows us to generate the inverse for any type ③ elementary matrix. Suppose E is a type ③ elementary matrix and $e_{i,j} = a$ is the off-diagonal, non-zero entry, and further suppose that F is the inverse of E. Then F is equal to E except for the one entry $f_{i,j} = -a$. An example of a type ③ elementary matrix and its inverse are given below.

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & 0 & 1 \end{pmatrix} \quad E^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -a & 0 & 1 \end{pmatrix}$$

From the previous discussion we can see that all elementary matrices have inverses. We can also calculate them.

Recall that a matrix A can be put in reduced echelon form by applying a finite series of row operations. Since multiplication by elementary matrices is the same as row operations, we can write

$$E_k E_{k-1} \cdots E_2 E_1 A = R$$

where R is the reduced echelon form of a matrix A and k is the number of row operations needed. This lets us prove the following theorem:

Theorem 3.1. *A square matrix A has an inverse if and only if its reduced echelon form R is an identity matrix.*

Proof. First suppose that A is invertible. Then we know that R is invertible. This comes from the equation $E_k E_{k-1} \cdots E_2 E_1 A = R$. Each of the elementary matrices E is invertible and we have the left side of the equation being the product of invertible matrices, which is also invertible (why?). Therefore, R is a square matrix that is invertible and in reduced echelon form. Such a matrix must be the identity matrix (since R is square and, by definition, the leading entries are equal to one).

Alternatively, suppose R is an identity matrix. This gives us

$$E_k E_{k-1} \cdots E_2 E_1 A = \mathbf{I}$$

Multiplying on the left of both sides of the equation, we have

$$E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} E_k E_{k-1} \cdots E_2 E_1 A = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} \mathbf{I}$$

Which collapses to

$$A = E_1^{-1}E_2^{-1}\cdots E_{k-1}^{-1}E_k^{-1}$$

The product on the right of the equation is the product of the inverses of elementary matrices and we can easily show that the inverse of an elementary matrix is also an elementary matrix. Therefore, the product on the right is itself an invertible matrix, so A is invertible and that

$$A^{-1} = E_kE_{k-1}\cdots E_2E_1$$

$\square$

This theorem gives us a recipe (or **algorithm**) that will let us calculate the inverse of of matrix because it shows that the same row operations that reduces a matrix A to $\mathbf{I}$ will, when applied to $\mathbf{I}$, change $\mathbf{I}$ into A^{-1}. It also shows that if A cannot be row reduced into $\mathbf{I}$, then A has no inverse.

The easy way to calculate the inverse of a matrix $A_{m\times m}$ is to create a $(m \times 2m)$ matrix such that the left half is the matrix A and the right half is $\mathbf{I}$. Then apply row operations to this matrix which will reduce A to $\mathbf{I}$. When this is done, the right half of the matrix will contain A^{-1}.

For example, let us find the inverse of A when

$$A = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 3 & 4 \\ 3 & 9 & 2 \end{pmatrix}$$

Write the augmented matrix $(A|\mathbf{I})$.

$$\left(\begin{array}{ccc|ccc} 2 & 4 & 8 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 3 & 9 & 2 & 0 & 0 & 1 \end{array}\right)$$

Interchange rows 1 and 2

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 4 & 0 & 1 & 0 \\ 2 & 4 & 8 & 1 & 0 & 0 \\ 3 & 9 & 2 & 0 & 0 & 1 \end{array}\right)$$

Add -2 times row 1 to row 2

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 4 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 & -2 & 0 \\ 3 & 9 & 2 & 0 & 0 & 1 \end{array}\right)$$

Add -3 times row 1 to row 3

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 4 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 & -2 & 0 \\ 0 & 0 & -10 & 0 & -3 & 1 \end{array}\right)$$

Multiply row 2 by $-\frac{1}{2}$

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 4 & 0 & 1 & 0 \\ 0 & 1 & 0 & -\frac{1}{2} & 1 & 0 \\ 0 & 0 & -10 & 0 & -3 & 1 \end{array}\right)$$

Multiply row 3 by $-\frac{1}{10}$

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 4 & 0 & 1 & 0 \\ 0 & 1 & 0 & -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 & 0 & \frac{3}{10} & -\frac{1}{10} \end{array}\right)$$

Add -3 times row 2 to row 1

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 4 & \frac{3}{2} & -2 & 0 \\ 0 & 1 & 0 & -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 & 0 & \frac{3}{10} & -\frac{1}{10} \end{array}\right)$$

Add -4 times row 3 to row 1

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{2} & -\frac{16}{5} & \frac{2}{5} \\ 0 & 1 & 0 & -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 & 0 & \frac{3}{10} & -\frac{1}{10} \end{array}\right)$$

This shows us that

$$A^{-1} = \begin{pmatrix} \frac{3}{2} & -\frac{16}{5} & \frac{2}{5} \\ -\frac{1}{2} & 1 & 0 \\ 0 & \frac{3}{10} & -\frac{1}{10} \end{pmatrix}$$

Problems 3.5

1. Prove that the product of invertible matrices is invertible.

2. Show that the inverse of an elementary matrix is also an elementary matrix.

3. The field $\mathbb{Z}_p$ is called **integers modulo a prime** , or integers mod p. For a given prime number p, $\mathbb{Z}_p$ consists of the integers $0, 1, \ldots, (p-1)$. For example, $\mathbb{Z}_3$ consists of the integers 0,1 and 2.

   For a field $\mathbb{Z}_p$, the binary operations of addition and multiplication are the same as for integers except that the result is defined to be **the remainder after the absolute value of the result is divided by p**. The number p is called the **modulus**. The remainder is called a **residue**. For example, in $\mathbb{Z}_3$, $2+2=1$ because 4 divided by 3 leaves a remainder (or residue) of 1. Similarly, $2 * 2 = 1$.[1]

[1]There are other ways of defining the integers that make up $\mathbb{Z}_p$.

What would be the additive inverses for the scalars that make up $\mathbb{Z}_3$? What would be the multiplicative inverses?

Why is the last digit of a non-zero square in $\mathbb{Z}_3$ always a 1? Why is the last non-zero digit a 2? (this was used, in 1952, by Robert Gauntt, then a freshman at Purdue University, to prove that $\sqrt{2}$ is irrational.)

Exercises 3.5

1. Multiply A and A^{-1} from our example.

2. Create the elementary matrices that correspond to the row operations of our example. Evaluate their product to show that it is equal to $\mathbf{I}$. Remember that matrix multiplication is not necessarily commutative so the order of multiplication is important.

3. Does the matrix
$$B = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 2 & 4 \\ 3 & 9 & 2 \end{pmatrix}$$
have an inverse? (**HINT**: Row reduce B to reduced echelon form.)

4. For the matrix P over the field $\mathbb{Z}_3$, find P^{-1}.
$$P = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & 2 & 2 \end{pmatrix}$$
(**HINT**: Use Lampp and select the correct field.)

5. Lampp can be used to examine what happens to lines and planes as we use row operations. Use Lampp to create the following rational matrix:
$$\begin{pmatrix} 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 \end{pmatrix}$$
Then graph this matrix. Note that two graphs are shown, each representing specific columns. Close the graph and use row operations to put this matrix in reduced row echelon form. Bring up the graph of the matrix after each row operation and observe the changes to the lines. Alternatively, one can scroll through the graphs by scrolling the operation button at the top right of the Lampp screen.

6. Create the following 3×8 matrix using Lampp.
$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Graph this matrix. Reduce the matrix to row reduced echelon form. View the graph of the matrix after each row operation. (**NOTE**: A plane which is on edge to the viewer is assumed to have zero thickness and is not drawn.)

7. Create the following 3×8 matrix using Lampp.

$$\begin{pmatrix} 1 & 1 & 0 & 8 & 1 & 0 & 0 & 8 \\ 1 & 0 & 1 & 8 & 0 & 1 & 0 & 8 \\ 0 & 1 & 1 & 8 & 0 & 0 & 1 & 8 \end{pmatrix}$$

Graph this matrix. Reduce the matrix to row reduced echelon form. View the graph of the matrix after each row operation.

8. Create the following 4×4 matrix using Lampp.

$$\begin{pmatrix} -1 & -1 & 1 & -10 \\ 1 & -1 & -1 & -10 \\ -1 & 1 & -1 & -10 \\ 1 & 1 & 1 & -10 \end{pmatrix}$$

Graph this matrix. Use the internet to find out the name of the 3-dimensional figure that is initially shown. Use the "/" button at the bottom left of the graph pane to zoom in. This changes the scales of the graph, starting first from (-10 to 10) to (-9 to 9). Try to think of what will be graphed by zooming using the "+" button to (-20 to 20).

9. Create the following 3×8 matrix using Lampp.

$$\begin{pmatrix} -1 & -1 & 1 & -10 & 1 & 0 & 0 & -10 \\ 1 & -1 & -1 & -10 & 0 & 1 & 0 & -10 \\ -1 & 1 & -1 & -10 & 0 & 0 & 1 & -10 \end{pmatrix}$$

Graph this matrix. Reduce the matrix to row reduced echelon form. View the graph of the matrix after each row operation. Change the scales of the graph.

Chapter 4

Applications

We now have enough tools to begin exploring how linear algebra is useful in a variety of areas. Since we are interested in applications, some mathematical theorems will be stated without proof. These theorems are not essential to an understanding of linear algebra.

4.1 Polynomials and Graphs

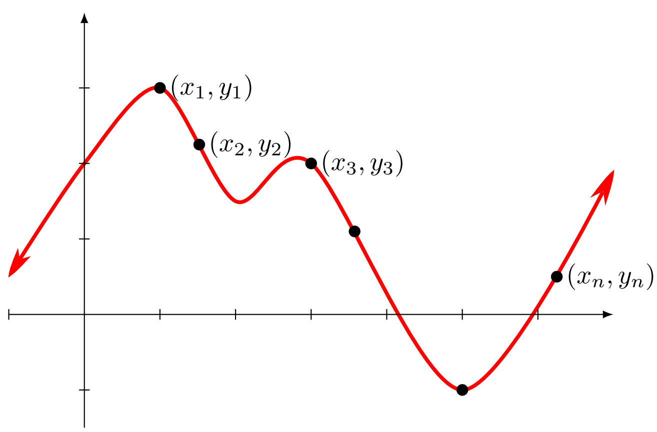


Figure 4.1: Part of a polynomial passing through n points.

A **polynomial** in a single variable is an equation of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

where $n \in \mathbb{Z} \geq 0$ and $a_i \in \mathbb{F}$ and $x^0 = 1$ for any $x \in \mathbb{F}$.[1] The **degree** of the

[1] We follow Donald Knuth's argument for $0^0 = 1$.

polynomial is defined to be n. In this text we will not discuss polynomials of more than one indeterminate variable.

Suppose we have n points in $\mathbb{R}^2$ (the xy-plane). Further suppose that we would like to fit a polynomial of degree $n-1$ that passes through all of these points, as illustrated in Figure 4.1.

In other words, we want to find a polynomial $p(x)$ such that

$$p(x) = \sum_{i=0}^{n-1} a_i x^i = y_i$$

We need to discover the n coefficients of $p(x)$ by substituting the x and y coordinates for the n points into the polynomial. This gives us n linear equations in n variables: $a_0, a_1, \ldots, a_{n-1}$.

$$\begin{array}{l}
a_0 + a_1 x_1 + a_2 x_1^2 \; + \cdots + a_n x_1^{(n-1)} = y_1 \\
a_0 + a_1 x_2 + a_2 x_2^2 \; + \cdots + a_n x_2^{(n-1)} = y_2 \\
\vdots \qquad\qquad\qquad\qquad \vdots \qquad\quad \vdots \\
a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_n x_n^{(n-1)} = y_n
\end{array}$$

We can then simply write this as an augmented $n \times (n-1)$ matrix and put it in reduced row echelon form. This allows us to read off the values of a_i.

We will follow with a numerical example where we determine the polynomial $p(x) = a_0 + a_1 + a_2 x^2$ whose graph passes through the points $(1,4)$, $(2,-1)$ and $(5,10)$. We begin by substituting our x and corresponding y values into a system of linear equations in the variables a_0, a_1 and a_2.

$$\begin{aligned}
p(1) &= a_0 + a_1(1) + a_2(1)^2 = a_0 + \;\; a_1 + \quad a_2 = \;\;\, 4 \\
p(2) &= a_0 + a_1(2) + a_2(2)^2 = a_0 + 2a_1 + \;\; 4a_2 = -1 \\
p(4) &= a_0 + a_1(5) + a_2(5)^2 = a_0 + 5a_1 + 25a_2 = \;\; 10
\end{aligned}$$

Solving, we get

$$a_0 = \frac{40}{3} \qquad\qquad a_1 = \frac{-21}{2} \qquad\qquad a_2 = \frac{11}{6}$$

so $p(x) = \frac{40}{3} - \frac{23}{2}x + \frac{13}{6}x^2$. Figure 4.2 shows this graph.

There are many types of curves besides polynomials which may be fitted to points on a graph. The preferred method is the one that gives the best approximation to points that are between given points (we say these points are **interpolated**) or points that are beyond all given points (**extrapolated** points).

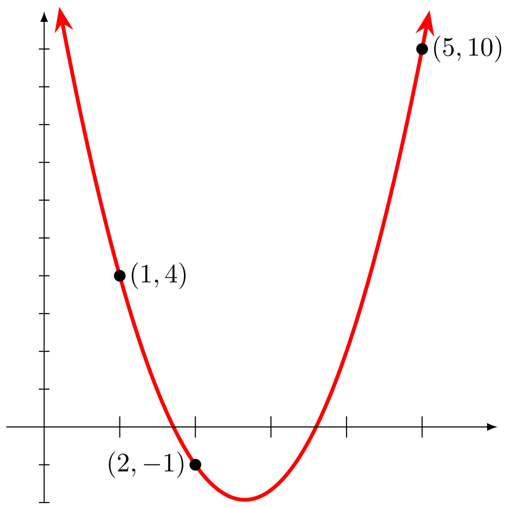


Figure 4.2: A polynomial passing through 3 points.

We will encounter another method of fitting points, called **linear regression**, in a later chapter.

Exercises 4.1

1. Choose three points from the following list and find a 2 degree polynomial to fit them (as in the example). Do the same for a different set of three points. What can you say about the two equations? (**Hint**: It is possible to plot polynomials using Lampp. Either middle click the mouse icon at the bottom right or down-swipe it. This will activate the polynomial plotter for the Graph panel. Select the use of columns by right-clicking or tapping the row-or-column toggle button near the top right. Then select the column of values you wish to use as the coefficients for the polynomial in a single indeterminate x you wish to graph.)

$$(1,2) \quad (2,5) \quad (4,2) \quad (6,2)$$

2. Fit all four points in the above example to a polynomial.

3. If the coordinate numbers of the points are too large for easy manipulation, it may be possible to transform them with a linear equation to make them smaller and more manageable. Suppose we have three points:

$$(10,-1) \quad (11,2) \quad (12,5)$$

We have x values of $\{10, 11, 12\}$. Set $z = x - 10$. Then we have the (z, y) points

$$(0,-1) \quad (1,2) \quad (2,5)$$

We find the polynomial

$$p(z) = \sum_{i=0}^{2} a_i z^i$$

then substitute $(x - 10)$ for all values of z.

$$\begin{aligned} p(z) &= a_0 + a_1 z + a_2 z^2 \\ p(x) &= a_0 + a_1(x - 10) + a_2(x - 10)^2 \end{aligned}$$

Find the polynomial coefficients for $p(x)$ directly. Then find $p(z)$ and use it to calculate $p(x)$. Demonstrate that the two forms of $p(x)$ are the same.

4.2 Graph Theory and Networks

Graph theory is a fairly new area of mathematics which is used to create models in economics, social sciences, traffic analysis, medicine, electrical engineering and many other fields. It has become a truly vast subject. We will try to give some flavor of it and show how the ideas of linear algebra are tightly connected.

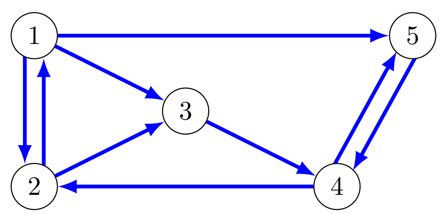


Figure 4.3: A Directed Graph

A **directed graph** is a collection of n points called **vertices** and a finite number of **edges**. Each pair of vertices is joined by one or more edges. Each edge is assigned a specific direction. In other words, if V_i and V_j are two vertices, then they can be connected by two edges in opposite directions, one going from V_i to V_j and one in the direction of V_i with its origin at V_j. Of course, two vertices can also be connected by many edges going in the same direction. An example of a directed graph is illustrated in Figure 4.3.

Any graph can be represented by a square matrix $A_{n\times n}$ such that $a_{i,j}$ is the number of edges joining vertex i to vertex j. The matrix representation for the directed graph illustrated in Figure 4.3 is given below.

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

Notice that the row number gives us the number of the vertex we come *from* and the column number corresponds to the vertex we travel *to*. This means that

$d_{1,5} = 1$ means there is an edge connecting vertex 1 with vertex 5. Since we cannot go directly to vertex 3 from vertex 5, we have $d_{5,3} = 0$.

A matrix representation of a directed graph D that has the conditions that $d_{i,j} = 0$ or 1 and that $d_{i,i} = 0$ is called a **incidence matrix**. This means that there is at most one edge from any vertex to any other. Also, no edge exists which connects a vertex to itself.

Incidence matrices can be used by sociologists and anthropologists to represent and study group interactions. These type of matrices can also be used to analyze transportation problems in exactly the same way. We will first look at a simple transportation system.

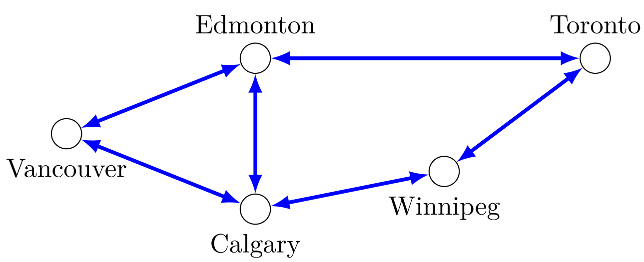


Figure 4.4: Example Airlines

In the sketch of the routes of our Example Airlines (Figure 4.4), we used a line with two arrowheads to represent when vertices are connected by edges in both directions. The cities could be given the following numbers for rows and columns:

Vancouver	1
Edmonton	2
Calgary	3
Winnipeg	4
Toronto	5

The incidence matrix for this diagram would be:

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

We can easily determine that it is impossible to make a direct flight from Vancouver to Winnipeg using Example Airlines. Intermediate stops are necessary. Suppose we wanted to know in how many ways we can travel from one city to another with a given number of stops. There are, for example, two routes a traveler can take from Vancouver to Winnipeg with two stopovers. A route from one vertex to another is called a **chain** or **path**. A path which is made up of exactly two edges is called a **2-chain**. An example would be a flight to Toronto from Calgary via Winnipeg. It should be very, very obvious what a 3-chain or an n-chain is composed of.

If we multiply our incidence matrix by itself, we get

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

We will see that $b_{i,j}$ gives us the number of ways we can go from city i to city j with exactly one stopover. First, notice that $a_{i,j}$ gives us the number of edges that connect vertex i with vertex j. In other words, there is one 1-chain that connects the two vertices.

We know from our definition of matrix multiplication that row 2 of A multiplied by column 4 of A gives us $b_{2,4}$.

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

The third column entry in row 2 of A says that there is a 1-chain from Edmonton (vertex 2) to Calgary (vertex 3). The row three entry of column 4 informs us that Calgary (vertex 3) is connected to Winnipeg (vertex 4). Multiplying the ones in these entries represents the 2-chain of Edmonton-Calgary-Winnipeg. Similarly, the multiplication of the last ones in the row and column represents the 2-chain Edmonton-Toronto-Winnipeg. This reasoning may be generalized and used to prove that A^m gives the number of ways to travel from city i to city j with exactly $m-1$ stopovers. We omit the proof but use the result to explore our graph further.

$$C = A^3 = \begin{pmatrix} 2 & 4 & 4 & 2 & 2 \\ 4 & 2 & 6 & 1 & 5 \\ 4 & 6 & 2 & 5 & 1 \\ 2 & 1 & 5 & 0 & 4 \\ 2 & 5 & 1 & 4 & 0 \end{pmatrix} \quad D = A^4 = \begin{pmatrix} 8 & 8 & 8 & 6 & 6 \\ 8 & 15 & 7 & 11 & 3 \\ 8 & 7 & 15 & 3 & 11 \\ 6 & 11 & 3 & 9 & 1 \\ 6 & 3 & 11 & 1 & 9 \end{pmatrix}$$

There are exactly three ways to fly Example Airlines from Toronto to Edmonton with three stopovers. They would be:

1. Toronto-Winnipeg-Calgary-Vancouver-Edmonton
2. Toronto-Edmonton-Calgary-Vancouver-Edmonton
3. Toronto-Edmonton-Vancouver-Calgary-Edmonton

Only the first 4-chain does not have Edmonton repeated. Any chain with a repeated vertex is called **redundant**.

Suppose an anthropologist has studied a clan of bonobos and observed the pattern of dominance given in Figure 4.5. Each V_i represents one member of the group.

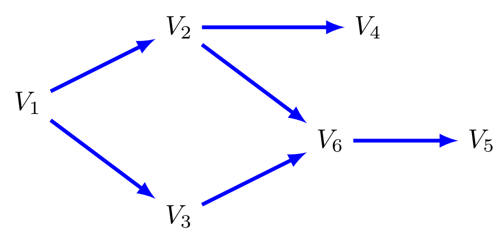


Figure 4.5: Dominance Patterns in a Group

We can see from this graph that the bonobo called V_1 directly dominates the bonobos V_2 and V_3 but only indirectly dominates the others. The incidence matrix and some of its powers for this directed graph would be:

$$V = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad V^2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$V^3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad V^4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Notice that in row 1 of V^2 there is a one in column 4 and a two in column 6. This indicates that there is one way that bonobo V_1 can dominate V_4 through an intermediary and two ways that V_1 can dominate V_6 through exactly one intermediary. Notice that these column positions in the matrix V are equal to zero. This means that V_1 does not directly dominate V_4 or V_6. There is no 1-chain connecting the two. Similarly we observe that the two in the matrix V^3 shows that there are two ways that bonobo V_1 can dominate V_5 with exactly two intermediaries. Notice that the row 1, column 5 entries in V and V^2 are zeros. This means that *the shortest path connecting* V_1 *and* V_5 *is a 2-chain.*

We can extend this logic to show that if we have an incidence matrix U such that U^n has a non-zero i, j entry, and if the i, j entry for every U^m is zero for $m < n$, then the shortest path from vertex i to vertex j is an $(m-1)$-chain. It should be apparent how useful this result can be.

One type of directed graph used frequently in economics, electrical engineering and traffic analysis is called a network. A **network** is a directed graph such that each pair of vertices is connected by at most one edge and each edge is assigned a number. The vertices are called **junctions** and the edges are called **branches**. The number assigned to each branch represents the **flow** along the branch. In most networks it is assumed that the total flow into a junction equals the total flow out of that junction. These are the type of networks we will handle since this property allows us to write the flow through a junction using linear equations. A network is represented by an augmented matrix since we need an extra column to deal with the numbers which represent the flow through each branch.

We illustrate how to assign linear equations with reference to Figure 4.6. Since the flow going into the junction, labeled x and y, is equal to the flow coming out of the junction, we can write the linear equation $x + y = 5$ to represent the flow through the junction.

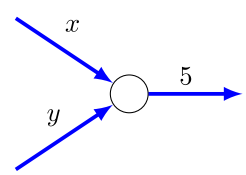


Figure 4.6: A Junction and 3 Branches

A more complicated network is given in Figure 4.7. In this network there are four junctions and nine branches. We can write the linear equations for each of

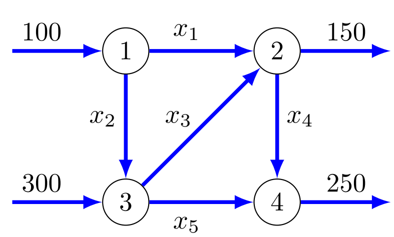


Figure 4.7: 4 Junctions and 9 Branches

the four junctions. The variables (and/or numbers) representing flow into the junction are put on the left of the equality sign. The variables for flow coming out are placed on the right.

$$\begin{aligned} 100 &= x_1 + x_2 && \text{Junction 1} \\ x_1 + x_3 &= 150 + x_4 && \text{Junction 2} \\ 300 + x_2 &= x_3 + x_5 && \text{Junction 3} \\ x_4 + x_5 &= 250 && \text{Junction 4} \end{aligned}$$

We use algebra to put these equations in a more traditional form.

$$\begin{aligned} x_1 + x_2 &= 100 && \text{Junction 1} \\ x_1 + x_3 - x_4 &= 150 && \text{Junction 2} \\ -x_2 + x_3 + x_5 &= 300 && \text{Junction 3} \\ x_4 + x_5 &= 250 && \text{Junction 4} \end{aligned}$$

We create an augmented matrix to hold our values.

$$\left(\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 0 & 100 \\ 1 & 0 & 1 & -1 & 0 & 150 \\ 0 & -1 & 1 & 0 & 1 & 300 \\ 0 & 0 & 0 & 1 & 1 & 250 \end{array}\right)$$

Then we use Gauss-Jordan to evaluate the matrix.

$$\left(\begin{array}{ccccc|r} 1 & 0 & 1 & 0 & 1 & 400 \\ 0 & 1 & -1 & 0 & -1 & -300 \\ 0 & 0 & 0 & 1 & 1 & 250 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

The rank of the coefficient matrix is 3 and since we have 5 variables we need 2 parameters. We'll call them s and t. This gives us a solution matrix:

$$\begin{pmatrix} 400 - s - t \\ -300 + s + t \\ s \\ 250 - t \\ t \end{pmatrix}$$

Notice that s and t cannot be any scalar $\in \mathbb{R}$. In order for the flow to follow the direction of the arrows, the values for each variable *must be positive*. So we must have $s > 0$ and $t > 0$. Also, from row 1 of the solution matrix, we must have $s + t < 400$. This will assure us a positive value for x_1. From row 2 we have that $s + t > 300$ for $x_2 > 0$ and from row 4 we have that $t < 250$. Putting all these inequalities together, we have

$$\begin{aligned} s &> 0 \\ t &> 0 \\ s + t &< 400 \\ s + t &> 300 \end{aligned}$$

We can then draw a graph to indicate the region containing all pairs (s, t) which give solutions which assure us of positive flow through the network. (Negative flow would be like having a toilet backing up, not a very desirable solution to a water flow problem.)

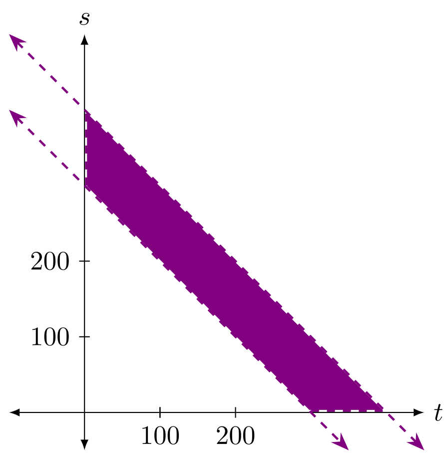
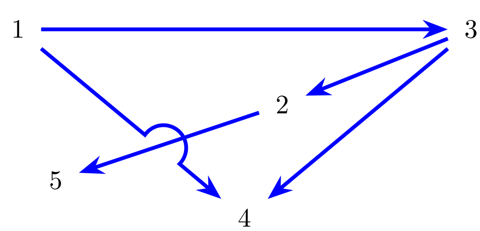


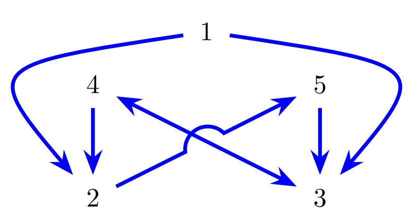
Figure 4.8: $s > 0, t > 0, s + t < 400$ and $s + t > 300$

Exercises 4.2

1. What would be the incidence matrix for the following diagram?



2. What would be the incidence matrix for the following sketch?



3. Sketch the diagrams for the following incidence matrices

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

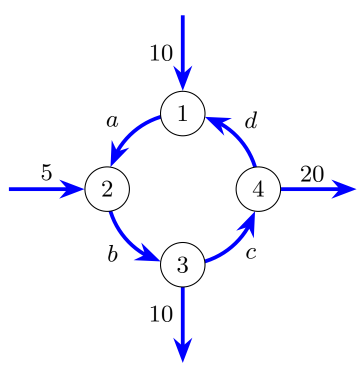
4. How many ways are there of traveling from vertex 3 to vertex 4 with exactly two stopovers for the directed graph given by the matrix A in the previous problem? (**HINT**: There is an undocumented (until now) feature in Lampp which allows for the quick multiplication of matrices. First, create a matrix A and populate it with the desired entries. Then press the B button on the bottom left. Accept the option to switch with A. Then create a matrix B and populate it with the desired entries. Now, swipe down (or middle-button-click) on the mouse icon at the bottom right corner. This activates the matrix Multiply options in the Matrix panel. Bring up the Matrix panel and select the button to multiply A with B and send the result to C. Note that the order of multiplication is given as $AB = C$. To get $BA = C$, you only need to select the appropriate button. After multiplication, the resultant matrix C is automatically displayed.) What is the shortest path from vertex 5 to vertex 4 in the directed graph represented by the incidence matrix B? What is this path?

5. Sketch the network for the following augmented matrix

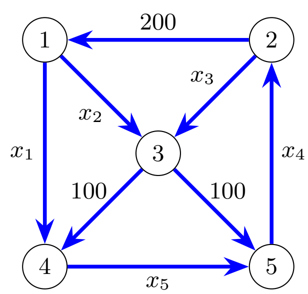
$$\left(\begin{array}{rrrr|r} 1 & -1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & -1 & 0 \\ 0 & -1 & -1 & 1 & -3 \end{array}\right)$$

6. Find all solutions for the network in the previous problem.

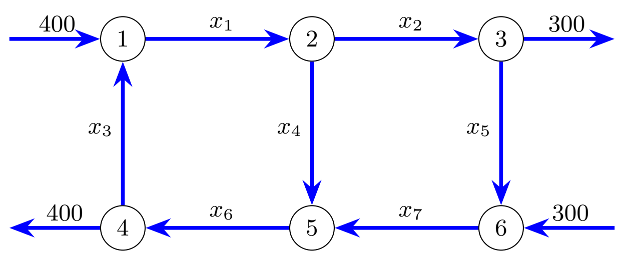
7. Solve for the network for the following diagram.



8. Find all solutions for the network for the following diagram.



9. Solve for the network for the following diagram.



4.3 Electrical Circuits

Analysis of electrical circuits is an extension of the ideas of networks of the previous section. We don't have to learn any new mathematics (*Alas!*), just the physical properties of electricity called **Kirchhoff's Laws**. These laws are usually stated as:

1. The current flowing into a junction equals the current flowing out of the junction.

2. The sum of the products of the current and resistance around a closed path is equal to the total voltage in the path.

We have already remarked that the networks we deal with already satisfy the first law. In order to make sense of the second law, we need to learn a little bit about electricity and a little more about flow.

Flow is a measure of how many things pass or how much of a thing passes through a branch or junction during a given interval of time. The number of cars per hour which travel along a road or pass through an intersection is a flow. The liters of water which makes its way through a tap in a second is a flow. The amount of stocks traded in a day, oil pumped through a pipeline in an hour, water down a river in a year, the number of library books checked out in a month, ..., flow,flow,flow,flow. The flow of electricity through a wire is called the **current**. It is measured in units called *amperes* (or amp).

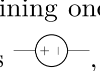
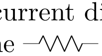
Flow can (and often does) change with time. Flow may vary in a cyclical way, as it does with electricity when we have *alternating current*. If the flow remains a constant (the rate never changes), we have *direct current*. The circuits we consider will be direct current networks.

A difference in pressure is usually what causes flow. In plumbing, a water pump can be the source of the pressure at one end of a pipe. In electrical circuits, a battery or a solar cell or something else may be the source of the pressure. This pressure is called the **voltage**, given in units called *volts* (usually abbreviated to the character V). Pressure in a direct current electrical circuit is applied in a direction. In electrical circuits we say that voltage is applied from the negative to the positive (from - to +). In other words, current will flow from the negative pole (cathode) of a battery to the positive pole (anode).[2]

Current can be inhibited completely if there are no wires connecting the cathode to the anode of a battery or any other voltage source. A **closed path** is a path which connects a vertex back to itself. A closed path is also called a **loop**. If the path is open, no current can flow. Current can be partially inhibited by a resistance to its flow. A **resistor** is a piece of material which converts some of the electricity to heat. Resistance is given in units called *ohms* (assigned the Greek omega symbol Ω). The voltage measured on each side of a resistor obviously cannot be the same. Since the resistor consumes some of the current, we say there is a *voltage drop* across the resistor. Each resistor in a circuit produces a voltage drop equal to the product of its resistance and the current that flows through it. We write this algebraically as

$$V = IR$$

where V is voltage, I is current and R is resistance. This formula is called Ohm's Law.

Figure 4.9 shows a closed circuit containing one voltage source and three resistors. The voltage sources are drawn as , with the current direction flowing from $-$ to $+$, and the resistors are each represented by the  symbol. The current flow is also shown by an arc in the clockwise direction and labeled I. We now have enough information to use Kirchhoff's second law to find the current and the voltage drop across each resistor.

We start by assigning variable names to each of the elements of the circuit. V would be the voltage induced by the battery so that $V = 45$ volts. Each resistor will be called R with a subscript indicating its rating in ohms, so that we have R_2,R_6 and R_7. I will stand for the current we will calculate, in amperes. From the second law we can write

$$V = IR_2 + IR_6 + IR_7$$

[2]The poles of batteries are also called terminals or nodes. The fact that we say current flows from the negative to the positive is an historical precedent set by Benjamin Franklin.

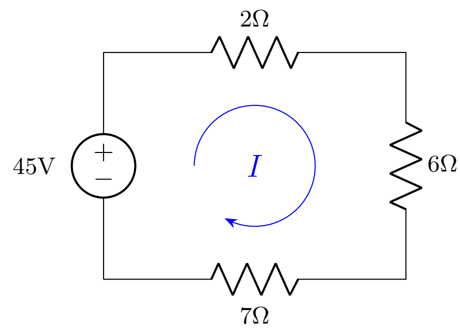


Figure 4.9: A Simple Closed Circuit

Plugging in the numbers gives us

$$45 = 2I + 6I + 7I = 15I$$

so $I = 3$ amp. We can then work back and show that the voltage drop across R_2, for instance, is $IR_2 = 3 * 2 = 18$ volts.

A more complicated system is given in Figure 4.10.

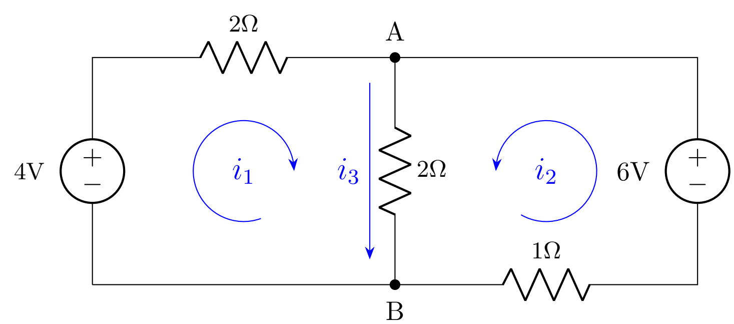


Figure 4.10: Two Voltage Sources and Three Resistors

To evaluate this circuit, we start by applying Kirchhoff's first law to junction A to get

$$i_1 + i_2 = i_3$$

or

$$i_1 + i_2 - i_3 = 0$$

Applying the first law to junction B will give the same result. We now use the second law to write equations for each of the loops. Starting with the left-hand loop, we can observe that the battery supplies 4 volts to the loop and the resistors have voltage drops of $2i_1$ and $2i_3$, since both resistances equal 2 ohms. Therefore we have

$$4 - 2i_1 - 2i_3 = 0$$

or

$$2i_1 + 2i_3 = 4$$

For the right-hand loop we have

$$i_2 + 2i_3 = 6$$

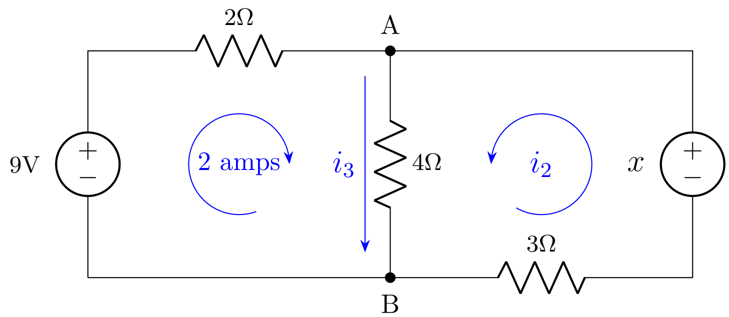
This gives us a system of linear equations

$$\begin{aligned} i_1 + i_2 - i_3 &= 0 \\ 2i_1 + 2i_3 &= 4 \\ i_2 + 2i_3 &= 6 \end{aligned}$$

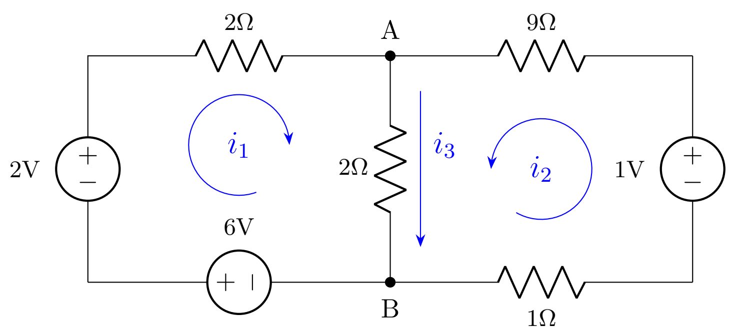
Using Gauss-Jordan we can then find that $i_1 = 2$ amp, $i_2 = 2$ amp and $i_3 = 2$ amp.

Exercises 4.3

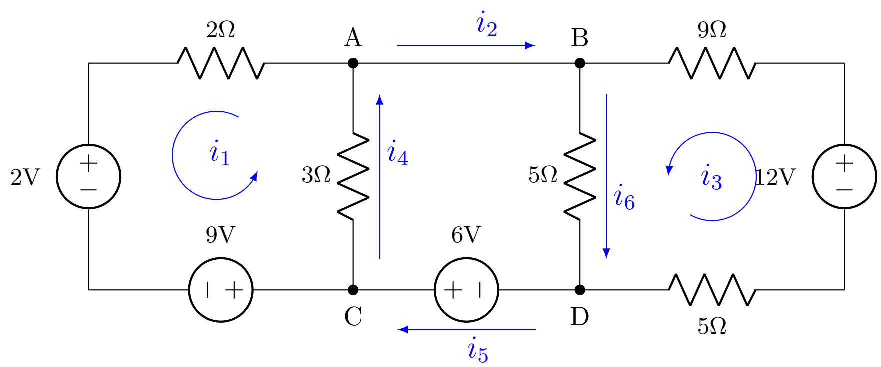
1. Find the voltage for the voltage source on the right. (Solve for x.)



2. Find the currents i_1, i_2 and i_3 for the following circuit.



3. Find values for all six currents for the following circuit. (**HINT**: When batteries are connected so that a positive terminal is connected directly to the negative terminal of the second battery, the result is like having one battery with a voltage value equal to the *sum* of the voltages of the two batteries. Connecting two nodes of the same sign would have the opposite effect.)



4.4 Cryptography

Cryptography is the art of encoding and decoding secret messages. The use of publicly accessible communication lines and the need to ensure information privacy has created a surge of interest in this area. In cryptography, coding algorithms are referred to as *ciphers*. Ciphers are used to change readable text (*plaintext*) to coded text (*ciphertext*). This process is called *enciphering* and the inverse action, creating plaintext from ciphertext, is called *deciphering*. One

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
P	Q	R	S	T	U	V	W	X	Y	Z	!	?	*	
15	16	17	18	19	20	21	22	23	24	25	26	27	28	

Table 4.1: A Simple Cipher

of the simplest ciphers is to replace letters of the alphabet with other letters or numbers. Suppose we had the correspondence of letters to numbers seen in Table 4.1.

Using this simple cipher, we can encode the message:

WHAT*IS*THE*FREQUENCY?

into

22 7 0 19 28 8 18 28 19 7 4 28 5 17 4 16 20 4 13 2 24 27

or, more likely, into

$$22070019280818281907042805170416200413022427$$

where each number is made two digits wide by preceding it with a zero if necessary.

This is the same procedure as performed by a computer when something is typed on the keyboard. The symbols on the keyboard are mapped onto unique numbers. The reverse happens when these numbers are sent to the monitor or printer for display.

This simple cipher is obviously too simple for serious secrecy. A more secure cipher system can be obtained by separating the plaintext into groups of n letters and replacing each group with n cipher letters. This is called a *polygraphic system*. We will look at one such method, which uses matrix multiplication, called *Hill ciphers*.

The most basic form of a Hill cipher involves breaking up a plaintext message into pairs of letters. An extra letter may be added to the plaintext to make an even number of characters. Our message would be grouped as

WH AT *I S* TH E* FR EQ UE NC Y?

which gives us the following pairs of numbers

$$22{,}7 \quad 0{,}19 \quad 28{,}8 \quad 18{,}28 \quad 19{,}7 \quad 4{,}28 \quad 5{,}17 \quad 4{,}16 \quad 20{,}4 \quad 13{,}2 \quad 24{,}27$$

We then choose an arbitrary 2×2 matrix of non-zero integer numbers *modulo 29*. The reason we choose to work in $\mathbb{Z}_{29}$ is because that is how many characters we need (and the field $\mathbb{Z}_{29}$ is available in Lampp). Suppose we choose the following matrix A

$$A = \begin{pmatrix} 7 & 2 \\ 3 & 10 \end{pmatrix}$$

We then write each pair of numbers as column matrices and multiply each matrix by A. For example, for the first pair of numbers we would have

$$\begin{pmatrix} 7 & 2 \\ 3 & 10 \end{pmatrix} \begin{pmatrix} 22 \\ 7 \end{pmatrix} = \begin{pmatrix} 23 \\ 20 \end{pmatrix}$$

We continue the process to get all the pairs of numbers for our ciphertext.

$$23{,}20 \quad 9{,}16 \quad 9{,}19 \quad 8{,}15 \quad 2{,}11 \quad 26{,}2 \quad 11{,}11 \quad 12{,}27 \quad 3{,}13 \quad 8{,}1 \quad 19{,}23$$

This corresponds to the following pairs of characters

XU JQ JT IP CL !C LL M? DN IB TX

which we would type up as

XUJQJTIPCL!CLLM?DNIBTX

and send off to our Fearless Leader at Secret Headquarters in Potsylvania.

To decipher the message, the ciphertext is broken up into pairs, translated into numbers and then each pair of numbers is multiplied by the inverse of the enciphering matrix A. Simply, $AP = C$ implies that $A^{-1}AP = A^{-1}C$, which gives us $P = A^{-1}C$, where A is the enciphering matrix, P is the matrix containing the pairs of plaintext numbers and C contains the pairs of ciphertext numbers. Of course, one must be careful that A *has* an inverse. We will look at this more closely in the chapter on determinants.

It is *fairly* easy to decipher a Hill cipher where the letters are grouped in pairs even if the encoding matrix is unknown (it would mean trying out 736 2×2 matrices over $\mathbb{Z}_{29}$). However, the difficulty of breaking a Hill cipher through trial-and-error increases enormously when larger groupings are used and the modulus is unknown.

Exercises 4.4

1. What is the inverse of the enciphering matrix A in our example of a Hill cipher. (**HINT**: Use Lampp and select the Integer Modulus a Prime option in the Field panel and enter a modulus of 29.)

2. Decode the following Hill ciphertext given that the enciphering matrix is

$$A = \begin{pmatrix} 12 & 16 \\ 25 & 7 \end{pmatrix}$$

OCVALDDG*GY?Y!OMD*KSRWREY*KI

Assume that the characters are grouped in pairs and encoded into numbers with the simple cipher of this section. (**HINT**: It is possible to use the MULTIPLY option in the MATRIX menu of Lampp to ease a lot of the computation. We can only multiply square matrices, so we need to group the cipher column matrix in pairs to form square, augmented matrices. If you don't know how to use the MULTIPLY option, read the exercises in the section on graph theory in this chapter. Better yet, do the exercises.)

4.5 Probability and Stochastic Matrices

When people assign a **probability** to an event, they are assigning a real number, p, such that $0 \le p \le 1$, and which represents the ratio of the number of ways the particular event can occur divided by the number of all possible events. When a coin is tossed, the probability is $p = \frac{1}{2}$ that a heads will show when the coin comes to rest. This number comes from the fact that there is only one way that a heads can show out of the total number of ways the coin will land (heads or tails). The same type of calculation gives us that the probability is $p = \frac{1}{12}$ that the sum of the numbers on the faces of a pair of thrown dice will be four. First we assume that the dice are thrown and we think of one die as the first die and the other as

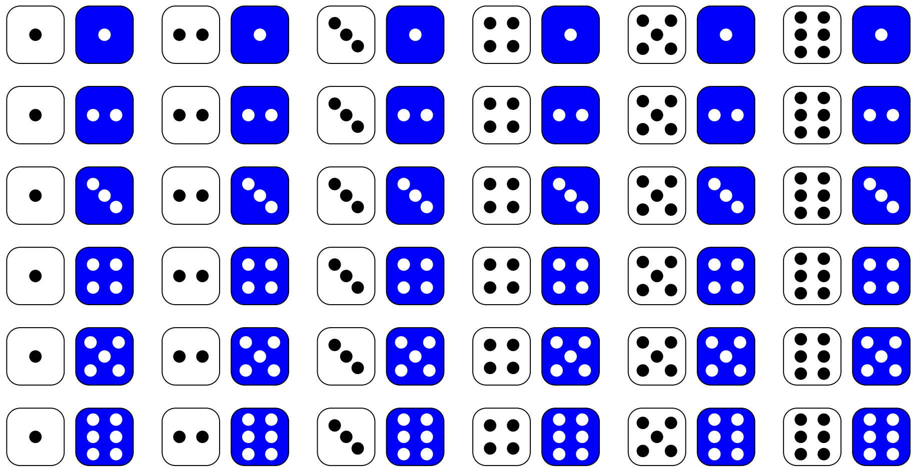


Figure 4.11: All possible outcomes of throwing a pair of dice

the second. There are six different ways the first die can fall. For each of these ways, there are six different faces the second die can show. Therefore there are $6 * 6 = 36$ different possible outcomes (possible events). There are three ways that numbers can show on the top faces so that they add up to four. The first die could have a 1 showing, which means the second must show 3. The second could have 1, so the first must show 3. Also, both could show 2. So we see that the probability is $\frac{3}{36} = \frac{1}{12}$ that the total on both faces is 4. This assumes that each outcome is as likely as any other.

If an event can never occur, the probability is 0. If the event can't fail to occur, the probability is 1. The probability of throwing 13 with a pair of dice is 0. The probability that the sum of the numbers is ≤ 12 on a pair of dice is 1. The probability that an event *will not* occur is one minus the probability that it *will* occur. The probability that a two and a four will show on a pair of thrown dice is $p = \frac{2}{36} = \frac{1}{18}$. The probability that we won't have a two and a four is $\neg p = 1 - p = \frac{17}{18}$. (The symbol $\neg$ is short for *not*.)

Probability is an estimate of real events. Though the probability is $\frac{1}{6}$ that a 5 will show when we throw one die, this does *not* mean that if we throw a die six times that a 5 will come up once. It *does* mean that it is reasonable to assume that a 5 will appear. It is a very good bet.

A **probability matrix** is a square matrix whose entries are probabilities and for which the sum of the entries in each row is 1. In other words, if $P_{n\times n}$ is a probability matrix, then

$$0 \leq p_{i,j} \leq 1 \quad \text{and} \quad \sum_{j=1}^{n} p_{i,j} = 1$$

A probability matrix is also called a **stochastic matrix**.[3]

Stochastic matrices have the nice property that they are closed under matrix multiplication. If we have that $A_{k\times k}$ and $B_{k\times k}$ are both stochastic matrices, then $C = AB$ is also a stochastic matrix. Let's prove this.

Proof. The sum of the entries in the mth row of C is

$$\sum_{j=1}^{k} c_{m,j} = \sum_{j=1}^{k}\sum_{i=1}^{k} a_{m,i} * b_{i,j} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{m,i} * b_{i,j}$$

$$= \sum_{i=1}^{k} a_{m,i} * \sum_{j=1}^{k} b_{i,j} = \sum_{i=1}^{k} a_{m,i} \quad \left[\text{because} \quad \sum_{j=1}^{k} b_{i,j} = 1\right]$$

$$= 1$$

□

One use for probability matrices is to describe phenomena which can have a finite number of different **states**

$$s_1, s_2, s_3, \ldots, s_n$$

any of which can be observed to occur at various *times* $t_0, t_1, t_2, \ldots$. For example, your blood pressure could be measured every hour and given one of the three ratings: high, normal or low. So your initial state, t_0, could be rated as high, the next hour could find your blood pressure unchanged (at $t_1 =$, your state is high) and then in the third hour your blood pressure state could be normal. We call such an arrangement of states and observation times a *system*. If the probability that a system is in a given state at a time t_i is *only* dependent on the state at time t_{i-1}, then the system is called a **Markov process**. Each path from an initial state to some other state is called a **Markov chain**. The state of such a system *does not* depend on any state previous to the one which immediately preceded it. Also, it does not occur because of the time itself. Thus, if your blood pressure state at one o'clock PM is high only because of its rating at noon, and not because it was high three hours previously or that it always shoots up after lunch, then your blood pressure system could be called a Markov process and your current blood pressure state could be considered the end of a Markov chain.

Let's make up an example of how a researcher might use a Markov process to predict future events. Suppose that an intrepid scientist discovered that there was an 80% chance that the child of a person who liked country music would also like country music. Also, there is a 15% chance that the child of a person who doesn't care for country music will spend countless hours swooning over Hank Williams, Sr., et al. What percentage of the population would like country music after 4 generations given that, initially, 40% of the population were tuning in to country music stations?

[3]The word stochastic comes from the Greek word for a person who predicts the future.

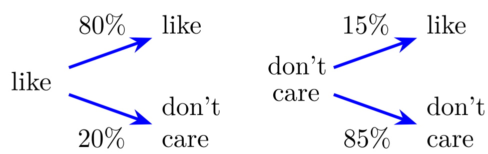


Figure 4.12: Tree Diagrams for a Markov Process

We start by drawing a special type of network called a **tree**. A tree is a network in which each junction has at most one branch with flow *into* the junction and at least one junction with no flow into it. For systems representing Markov processes, we are dealing with n possible states for each junction, so any junction will have *either* n branches flowing to other junctions *or no* branches flowing to other junctions. The tree diagrams for our country music system are given in Figure 4.12.

We use these diagrams to create the **transition matrix** for the Markov process. This will allow us to calculate each Markov chain. We first convert the percentages to probabilities (so that $80\% = \frac{4}{5}$, for instance). Then these probabilities are used to create the columns for the transition matrix.

$$T = \begin{pmatrix} \frac{80}{100} & \frac{15}{100} \\ \frac{20}{100} & \frac{85}{100} \end{pmatrix} = \begin{pmatrix} \frac{4}{5} & \frac{3}{20} \\ \frac{1}{5} & \frac{17}{20} \end{pmatrix}$$

The transition matrix T can now be used to calculate the next state for any given state. For example, given that initially 40% of the population likes country music, the next generation would have

$$\begin{pmatrix} \frac{4}{5} & \frac{3}{20} \\ \frac{1}{5} & \frac{17}{20} \end{pmatrix} \begin{pmatrix} \frac{2}{5} \\ \frac{3}{5} \end{pmatrix} = \begin{pmatrix} \frac{41}{100} \\ \frac{59}{100} \end{pmatrix}$$

From this we conclude that 41% of the next generation will like country music and 59% will be ambivalent. To calculate the state of the Markov process after n generations, we simply multiply the column matrix containing the initial state probabilities by T^n. For four generations of country music listeners we would have the transition matrix

$$T^4 = \begin{pmatrix} \frac{21223}{40000} & \frac{56331}{160000} \\ \frac{18777}{40000} & \frac{103669}{160000} \end{pmatrix}$$

We use this transition matrix to calculate the percentage of great-great-grandchildren who will be enjoying country music.

$$\begin{pmatrix} \frac{21223}{40000} & \frac{56331}{160000} \\ \frac{18777}{40000} & \frac{103669}{160000} \end{pmatrix} \begin{pmatrix} \frac{2}{5} \\ \frac{3}{5} \end{pmatrix} = \begin{pmatrix} \frac{338777}{800000} \\ \frac{461223}{800000} \end{pmatrix}$$

This shows that roughly 42% of the population will like country music after four generations. We would conclude that the percentage of the population who listens to country music remains fairly constant from generation to generation.

Exercises 4.5

1. Determine if the following matrices are stochastic.

   (a) $\begin{pmatrix} 1 & \frac{4}{5} \\ 0 & \frac{1}{5} \end{pmatrix}$ (b) $\begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$

   (c) $\begin{pmatrix} \frac{4}{10} & \frac{5}{10} & \frac{1}{10} \\ \frac{3}{10} & \frac{4}{10} & 0 \\ \frac{3}{10} & \frac{2}{10} & \frac{9}{10} \end{pmatrix}$ (d) $\begin{pmatrix} \frac{1}{12} & 0 & \frac{5}{24} \\ \frac{1}{4} & \frac{1}{2} & \frac{7}{24} \\ \frac{7}{12} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

2. Suppose a census indicated that 80% of the children of people with a university education will go on to university and that 10% of the children of people without a university education will obtain a degree. Represent the possible transitions between states and their probabilities with tree diagrams. Determine the transition matrix. Assuming that this trend remains constant between generations and that initially 30% of the population is university-educated, what percentage of the population will not have a university degree after 3 generations?

3. Repeat the previous exercise assuming that the initial population has 50% with university education.

4. A taxi company serves passengers in a city divided into three zones. Records indicate that of the passengers picked up in zone 1, half request to be taken to destinations in zone 1, 30% wish to go to zone 2 and the rest ask to be taken to zone 3. Of the passengers picked up in zone 2, 40% will go to zone 1, 20% to zone 2 and 40% wish transportation to zone 3. Of the passsengers picked up in zone 3, 20% will go to zone 1, 60% will go to zone 2 and 20% will be let off in zone 3. At the beginning of the day, 60% of the taxis are in zone 1, 10% in zone 2 and the rest are in zone 3 (30%). After all the taxis have had one rider, what is the distribution of taxis in the three zones? After they all have transported two riders?

4.6 Leontif Economic Models

A mathematical **model** is a set of equations which is used to mimic a real-world system. A model is considered accurate or successful if it can be used to predict the behavior of a system as situations change. The set of equations we use to determine voltages in an electrical circuit is one such model. It is considered very successful as it approximates voltage measurements in real circuits very closely and the variables which primarily affect the system are few and well understood. Models in economics are rarely so successful because the number of things which can affect a market system are enormous and inter-relationships can be very complex. Even simple models, though, can be useful to provide indicators of behavior of an economic system.

The Leontif Open Production Model (often referred to as Input-Output Analysis) can be used to describe gross relationships between various industrial sectors. In particular, it can be used to analyze how much output each sector must produce in order to meet consumption and export demands.

We start by dividing up an economy into a number of industries. For example, we might have transportation, steel, agriculture, entertainment, etc. Each industry uses resources which we call inputs. The outputs from each sector are the services or manufactured goods provided by the people and technologies associated with the industry. The inputs to each industry are made up in part by the outputs of other sectors. The agriculture industry, for example, uses the output of many industries, such as transportation (trucks, tractors), oil (fertilizers, motor fuel) and lumber (wood). We use an *input-output matrix* to summarize the interdependence among the industries we are interested in studying. An input-output matrix typically looks like:

$$\begin{array}{c c} & \begin{array}{cccc} \text{Sector 1} & \text{Sector 2} & \dots & \text{Sector } n \end{array} \\ \begin{array}{c} \text{Sector 1} \\ \text{Sector 2} \\ \vdots \\ \text{Sector } n \end{array} & \left(\begin{array}{cccc} \phantom{\text{Sector 1}} & \phantom{\text{Sector 2}} & & \phantom{\text{Sector } n} \\ \\ \\ \\ \end{array} \right) \end{array}$$

The i, j entry of the matrix gives us the output from Sector i that is used as input to Sector j. It is usually most convenient to express the entries in monetary units, so that a column lists the dollar values of every input the sector requires to produce \$1 of output.

The model must also take into account consumers who are not members of the represented industries. These consumer requirements are called the *demand* on the economy. This demand is represented by a column matrix indicating the amount required from each industry:

$$\text{demand} = \begin{pmatrix} \text{amount from Sector 1} \\ \text{amount from Sector 2} \\ \vdots \\ \text{amount from Sector } n \end{pmatrix}$$

So let us suppose that an economy was composed of only three industries. We will say these are manufacturing, agriculture and energy. We will say that it takes \$0.20 worth of manufactured goods, \$0.25 worth of energy production and \$0.05 of food inputs in order to produce \$1.00 of agricultural produce. Furthermore, the manufacturing sector requires \$0.15 of agricultural produce, \$0.15 of its own product and \$0.30 worth of energy to make \$1.00 of manufactured goods. Finally, we stipulate that the energy industry requires \$0.05 of agricultural output and \$0.10 worth of manufactured products to stay in business and that it only consumes \$0.01 of its own output. How much should each sector produce in order to meet the demand for exports of \$2 billion of agricultural produce, \$1 billion of manufactured goods and \$3 billion worth of energy exports?

First, we set up the input-output matrix A

$$\begin{array}{l} \\ \text{Agriculture} \\ \text{Manufacturing} \\ \text{Energy} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Agriculture} & \text{Manufacturing} & \text{Energy} \end{array} \\ \left(\begin{array}{ccc} .05 & .15 & .05 \\ .20 & .15 & .10 \\ .25 & .30 & .01 \end{array}\right) \end{array} = A$$

We let D be the demand matrix (in billions of dollars)

$$D = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

We let x_1 = the amount the agricultural sector produces, x_2 = the value of the manufactured goods created and x_3 = the value of the energy created by the energy sector. We wish to know how much of each of these values is left after all of the industries satisfy their requirements. In other words, we need

$$\begin{aligned} &x_1 - [\text{agricultural goods used in production}] \\ &x_2 - [\text{manufactured goods used in production}] \\ &x_3 - [\text{energy used in production}] \end{aligned}$$

We then determine formulas for the amounts used in production

$$\begin{aligned} [\text{agricultural goods used in production}] &= .05x_1 \ + \ .20x_2 \ + \ .25x_3 \\ [\text{manufactured goods used in production}] &= .15x_1 \ + \ .15x_2 \ + \ .30x_3 \\ [\text{energy used in production}] &= .05x_1 \ + \ .10x_2 \ + \ .01x_3 \end{aligned}$$

If we let

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

then what we are asking for can be expressed simply as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} .05 & .15 & .05 \\ .20 & .15 & .10 \\ .25 & .30 & .01 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = X - AX$$

We then equate this with our demand to get

$$X - AX = D$$

To solve this equation, we substitute $\mathbf{I}X$ for X and then do a little manipulation

$$\mathbf{I}X - AX = D$$

$$(\mathbf{I} - A)X = D$$

which finally gives

$$X = (\mathbf{I} - A)^{-1}D$$

Applying arithmetic to our example then gives us

$$\mathbf{I} - A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} .05 & .15 & .05 \\ .20 & .15 & .10 \\ .25 & .30 & .01 \end{pmatrix} = \begin{pmatrix} .95 & -.15 & -.05 \\ -.20 & .85 & -.10 \\ -.25 & -.30 & .99 \end{pmatrix}$$

Applying Gauss-Jordan elimination gives us

$$(\mathbf{I} - A)^{-1} = \frac{1}{14477}\begin{pmatrix} 16230 & 3270 & 1150 \\ 4460 & 18560 & 2100 \\ 5450 & 6450 & 15550 \end{pmatrix}$$

Solving for our demand matrix D gives us

$$X = (\mathbf{I} - A)^{-1}D =$$

$$\frac{1}{14477}\begin{pmatrix} 16230 & 3270 & 1150 \\ 4460 & 18560 & 2100 \\ 5450 & 6450 & 15550 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{14477}\begin{pmatrix} 39180 \\ 33780 \\ 64000 \end{pmatrix}$$

We round the results to two decimal places to get

$$X = \begin{pmatrix} 2.71 \\ 2.33 \\ 4.42 \end{pmatrix}$$

In other words, we need the agriculture sector to produce \$2.71 billion worth of produce, manufacturing \$2.33 billion, and energy should output \$4.42 billion in order to meet the demand.

Exercises 4.6

1. Suppose a corporation has three divisions; a plastics division, a controllers division and a semiconductor division. For each dollar of production, the plastics division requires \$0.10 of its own product, \$0.20 worth of controllers and \$0.10 worth of semiconductors. The controllers division requires no plastics, \$0.10 worth of controllers and \$0.20 worth of semiconductors. The semiconductor division uses \$0.10 worth of plastics and \$0.10 of controllers. It does not use any of its own output. How much should each division produce if there is a demand for \$10 million of plastics, \$20 million of controllers and \$5 million of semiconductors?

2. Suppose, for the previous problem, the demand for semiconductors doubled and the demand for controllers dropped by 50%. What should the production for each division be?

3. Create an economy with four sectors and solve it for three different demand matrices.

4.7 Chemical Reactions (Stoichiometry)

Chemistry is the branch of science concerning the composition, structure and properties of matter and the transformations that matter undergoes. Stoichiometry (pronounced *stoy-kee-OM-uh-tree*) is the branch of chemistry which deals with the quantitative relationships derived from chemical equations and formulas. Ordinary matter is composed of atoms which are conserved in chemical reactions. Atoms may join together to form particles called molecules which share certain properties. Chemical reactions often involve the transformation of different types of molecules. Since the number of atoms are conserved, chemists express this equality as balanced equations. An example might be:

$$2\mathrm{Na} + 2\mathrm{H_2O} \rightarrow 2\mathrm{NaOH} + \mathrm{H_2}$$

The balanced equation states that 2 molecules of sodium (Na) mixed with two molecules of water (H_2O) will produce 2 molecules of sodium hydroxide (NaOH) and one molecule of hydrogen gas (H_2). One molecule of sodium contains one atom of sodium. One molecule of water is composed of 2 atoms of hydrogen (H) for every atom of oxygen (O). The white, caustic sodium hydroxide is made up of one atom each of sodium, oxygen and hydrogen. Hydrogen molecules are usually composed of two atoms of hydrogen. We can therefore write the balanced equation as a system of linear equations.

$$\begin{array}{lccccc}
 & 2\mathrm{Na} & +2\mathrm{H_2O} & \rightarrow & 2\mathrm{NaOH} & + \mathrm{H_2} \\
\mathrm{Na} & 2 & & = & 2 & \\
\mathrm{H} & & 4 & = & 2 & + \ 2 \\
\mathrm{O} & & 2 & = & 2 &
\end{array}$$

Or, using familiar notation:

$$\begin{aligned}
2x_1 &= 2x_3 \\
4x_2 &= 2x_3 + 2x_4 \\
2x_2 &= 2x_4
\end{aligned}$$

It should be obvious which values of x_i correspond to which number of molecules.

Usually we have an unbalanced equation which we would like to put into a balanced form. For example, given the following unbalanced equation:

$$\mathrm{Cl_2} + \mathrm{NaBr} \rightarrow \mathrm{Br_2} + \mathrm{NaCl}$$

To balance this equation, we assign appropriate x_i variables and write out the system of linear equations. Then we solve the system and write out the balanced equation. We would start with:

$$x_1\mathrm{Cl_2} + x_2\mathrm{NaBr} \rightarrow x_3\mathrm{Br_2} + x_4\mathrm{NaCl}$$

which gives us

$$\begin{aligned} \text{chlorine (Cl)} \quad 2x_1 &= x_4 \\ \text{sodium (Na)} \quad x_2 &= x_4 \\ \text{bromine (Br)} \quad x_2 &= 2x_3 \end{aligned}$$

which can also be written:

$$\begin{aligned} 2x_1 - x_4 &= 0 \\ x_2 - x_4 &= 0 \\ x_2 - 2x_3 &= 0 \end{aligned}$$

where x_1 is the number of molecules of chlorine, x_2 is the number of molecules of sodium bromide, x_3 is the molecules of bromine and x_4 is the molecules of sodium chloride (salt).

We write this as an array and solve it using Gauss-Jordan elimination.

$$\begin{pmatrix} 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & -2 & 0 \end{pmatrix} \text{ which reduces to } \begin{pmatrix} 1 & 0 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -\frac{1}{2} \end{pmatrix}$$

Using a parameter, we let $t = x_4$ to get $x_3 = 2t$, $x_2 = t$ and $x_1 = t$. Since t is arbitrary, we assign $t = 2$ and get that $x_1 = 1$, $x_2 = 2$, $x_3 = 1$ and $x_4 = 2$. This means our balanced equation would be:

$$Cl_2 + 2NaBr \to Br_2 + 2NaCl$$

Exercises 4.7

Balance the given chemical equations for the following exercises.

1. $CH_4 + O_2 \to CO_2 + H_2O$
2. $Al_2(SO_4)_3 \to AlCl_3 + CaSO_4$
3. $ZnS + O_2 \to ZnO + SO_2$
4. $Ba(OH)_2 + CO_2 \to BaCO_3 + H_2O$
5. $HI + KMnO_4 + H_2SO_4 \to I_2 + MnSO_4 + K_2SO_4 + H_2O$

Chapter 5

Vector Spaces

In order to continue our exploration of applications, we must first return to more abstract reasoning. Readers should consider these next few chapters as a break from real-world problems, a vacation where we are engaging in a game with symbols and well-defined rules. Later on we will see how this will benefit in arriving at solutions to concrete problems.

5.1 Vectors and Plane Geometry

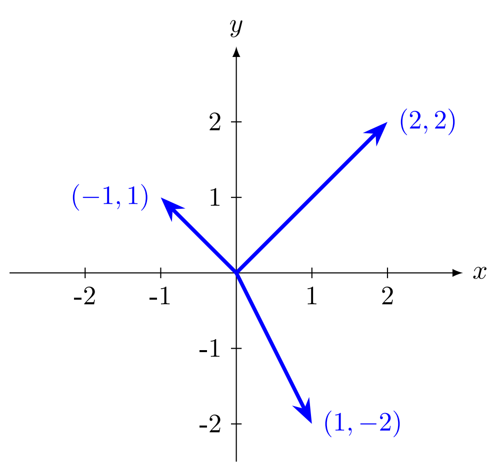


Figure 5.1: Graphical Examples of Vectors

As was stated in the preface, a **vector** can be thought of as an ordered list of scalars from a field.[1] This is one type of vector. Another type of vector can be

[1] There are spaces composed of ordered lists of infinite numbers of variables. These figure prominently in modern physics, but not in this introductory text.

thought of as something with a magnitude and a direction, like a directed line segment (an arrow) in 3-dimensions. We will use both of these representations but we will consider a more general definition that defines a vector by its membership in a type of set.

Definition 5.1. *A **vector space** (also called a **linear space**) $\mathcal{V}$ over a field $\mathbb{F}$ is a set composed of elements such that:*

1. *For any three elements $\vec{u}, \vec{v}, \vec{w} \in \mathcal{V}$, there is a binary composition called vector addition which satisfies:*
   - *(a)* $\vec{v} + \vec{w} \in \mathcal{V}$
   - *(b)* $\vec{v} + \vec{w} = \vec{w} + \vec{v}$
   - *(c)* $\vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$
   - *(d) there is a unique* $\vec{0} \in \mathcal{V}$ *such that* $\vec{0} + \vec{v} = \vec{v} + \vec{0} = \vec{v}$
   - *(e) there is a unique element* $-\vec{v} \in \mathcal{V}$ *such that* $\vec{v} + (-\vec{v}) = \vec{0}$
2. *For any scalars $a, b \in \mathbb{F}$ and any $\vec{v}, \vec{w} \in \mathcal{V}$, there is a scalar product $a\vec{v}$ which satisfies:*
   - *(a)* $a\vec{v} \in \mathcal{V}$
   - *(b)* $a(\vec{v} + \vec{w}) = a\vec{v} + a\vec{w}$
   - *(c)* $(a + b)\vec{v} = a\vec{v} + b\vec{v}$
   - *(d)* $(ab)\vec{v} = a(b\vec{v})$
   - *(e)* $1\vec{v} = \vec{v}$

A **vector** is an element of a vector space. It is defined by how it adds to other vectors, how it is multiplied by scalars and that it has an additive inverse. Nothing is mentioned about the location of a vector or how we would go about multiplying two vectors. By this definition of vector space we may only compare the relative size of two vectors if they are scalar multiples of each other. There is nothing to tell us of the size of a vector in relation to vectors that are not scalar multiples. The abstract nature of this definition assures us that theorems created for vectors are very general and will apply in many scenarios but this also makes it difficult to give a geometric interpretation of vectors. To do so, we will enrich the definition of a vector space with a few more ideas in order to make a connection between Euclidean geometry and vector algebra. First, we use the above definition to distinguish between two common geometric representations of vectors $\in \mathbb{R}^n$.

We say that, for instance, **two-dimensional Euclidean space** is the set of all ordered pairs (x_1, x_2) of real numbers, along with the traditional addition and multiplication operations. We refer to this space as $\mathbb{R}^2$ and usually identify it geometrically as a plane, with each ordered pair considered a **point** on that plane.

For our purposes, a point on a plane will be a identified by a lowercase, italicized character and be composed of an associated list of scalars. For example,

$p = (1, 0)$ and $q = (-5, 2)$ would be two points in $\mathbb{R}^2$. We will use a lowercase, italicized character with a diacritical arrow like $\vec{a}$ to refer to a vector. We can then write $\vec{a} = (-1, 1)$, $\vec{u} = (2, 3)$, etc. Notice that points and vectors, in this representation, are identical, *if the addition of points and the multiplication of points by scalars is taken to be the same as for vector space elements.* In a later chapter, we will augment a vector space in order to define a difference between a point and a vector. We will make a distinction by identifying two different types of geometrical vectors. We define one type immediately below and another type when we discuss the geometric product.

Definition 5.2. *A* ***free vector*** *is a directed line segment anywhere in a space.*

One can think of a free vector as an arrow that always points in the same direction but has *no* fixed position. The following figure 5.2 is of **only 3 different free vectors**, because the arrows of the same length that point in the same direction, the three arrows on the left, are drawings of the same vector. The fourth arrow points in the same direction as the leftmost three but has a relatively different length. It, too, represents a vector, one that is a non-zero and non-one, scalar multiple of the first vector. The rightmost arrow is an image of a third, different vector, as it points in a different direction. The two endpoints of a free

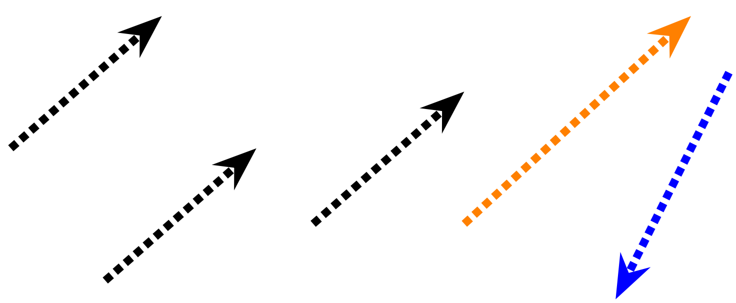


Figure 5.2: Five Images of Three Different Free Vectors

vector $\in \mathbb{R}^n$ are called, respectively, an initial point, or **tail**, and a final point, or **head**. The tail of a free vector may be any point in a space but the head is a point always in a certain direction from the vector's tail. Think of a free vector as a matrix of two columns, the first column being the tail and the second, the head. It might look like:

$$\vec{f} = \begin{pmatrix} t_1 & c_1 - t_1 \\ t_2 & c_2 - t_2 \\ \vdots & \vdots \\ t_n & c_n - t_n \end{pmatrix}$$

where t_i are variables representing any scalars $\in \mathbb{F}$ and c_i are assigned scalars $\in \mathbb{F}$. Figure 5.3 shows one free vector $\vec{f}$ and some possible positions that it may occupy on the $\mathbb{R}^2$ plane. (In reality, nobody writes free vectors this way anymore. They write them, most economically, as lists just using the c_i's.)

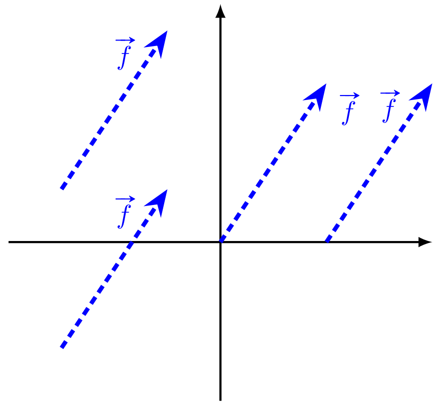


Figure 5.3: Another Example of a Free Vector

Confusion may arise because the term vector is used differently in different professions. To an epidemiologist, a vector is the name given to the carrier of a disease-causing agent that moves the agent to different areas on the planet. A vector, mathematically, was originally conceived as a method that 'carried' points between two positions in a geometrical space. Nowadays a vector, to a mathematician, is a very general object and could even be a matrix or a function, but it is always an element of a vector space. To a physicist or engineer, a vector is a physical quantity which has a direction and a magnitude, so it is often drawn as an arrow on a plane or in three dimensions. However, a physicist is usually also interested in the point of application of a force so the directed line segment (arrow) which represents the quantity likely does not emanate from an origin. Physicists and engineers almost always use free vectors in their models of physical systems, as physicals laws do not change with time or location. Then, when calculating the effects of applied forces for specific instances, they (explicitly or implicitly) define an origin and use directed line segments that emanate from the origin as vectors to calculate a result.

Unless explicitly stated otherwise, all references to vectors in this text mean elements of a vector space.

Two matrix forms of vectors are **column vectors** and **row vectors**. A row vector would be listed like a $1 \times m$ matrix, so $\vec{u} = (u_1, u_2, \ldots, u_m)$ would be an example of a row vector. Similarly, an $n \times 1$ matrix of the form

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

could be considered a column vector if it is also an element of a vector space.

Committing the properties of a vector space to memory is greatly facilitated by examining vectors in $\mathbb{R}^2$ over $\mathbb{R}$. (Proving that $\mathbb{R}^2$ is a vector space is given as a problem at the end of this section. We will take it on faith for the moment.)

Now let us look at a geometric interpretation of vector addition.

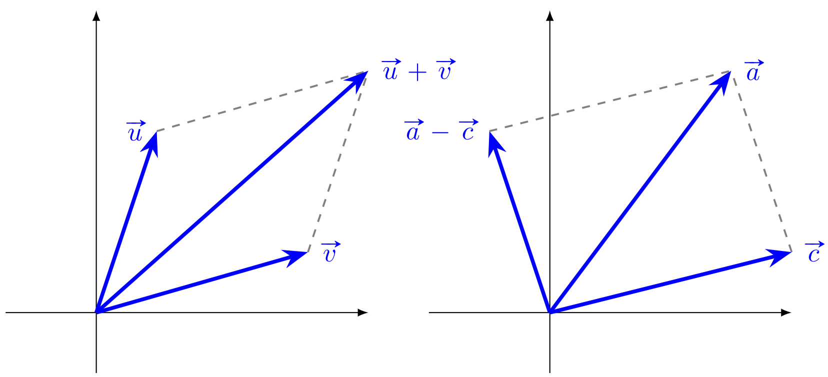


Figure 5.4: Vector Addition and Subtraction

If there are two vectors $\vec{u} = (x_1, y_1)$ and $\vec{v} = (x_2, y_2) \in \mathbb{R}^2$, then $\vec{u}+\vec{v} = \vec{w}$ is defined as $\vec{w} = (x_1+x_2, y_1+y_2)$. Also, given $\vec{a} = (x_a, y_a)$ and $\vec{c} = (x_c, y_c) \in \mathbb{R}^2$, then $\vec{a} - \vec{c} = (x_a - x_c, y_a - y_c)$. This is illustrated in Figure 5.4. If we have a directed line segment which starts at the point (x_1, y_1) and has its terminal point at (x_2, y_2), then a vector which has the same magnitude and direction and which emanates from the origin can be given as $(x_2 - x_1, y_2 - y_1)$. In this way we **translate** the directed line segment to the origin. Any directed line segment which can be translated to a given vector is said to be **equivalent** to that vector. A free vector is equivalent to a unique directed line segment emanating from the origin. We can see in Figure 5.4 that the directed line segment from the terminal point of $\vec{u}$ to the terminal point of $\vec{u} + \vec{v} = \vec{w}$ is equivalent to $\vec{v}$. Similarly, the directed line segment from the terminal point of $\vec{a} - \vec{c}$ to the terminal point of $\vec{a}$ is equivalent to $\vec{c}$.

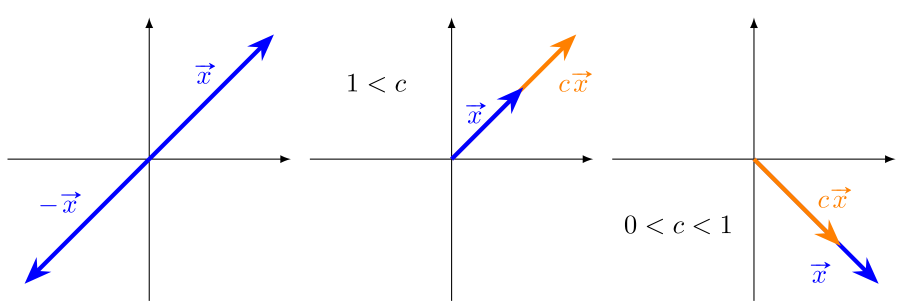


Figure 5.5: Scalar Multiplication of Vectors

Multiplying a non-zero vector by -1 will give a vector of the same magnitude which points in the opposite direction of the original vector. Multiplying a non-zero vector by a positive scalar will expand or compress the original vector, depending if the scalar is greater than one or less than one, respectively. The vector is *scaled*, hence the term *scalar*. Figure 5.5 shows these three scalar multiplications.

A vector $\in \mathbb{R}^2$ can be said to have 2 **components**. In the case of a vector (x, y), the components are the two real numbers x and y.

We have mentioned directed line segments, the direction of arrows and the magnitude of a vector but have not been precise by what we mean. We will review a little **trigonometry** and then give a precise definition of the magnitude and direction of a vector.

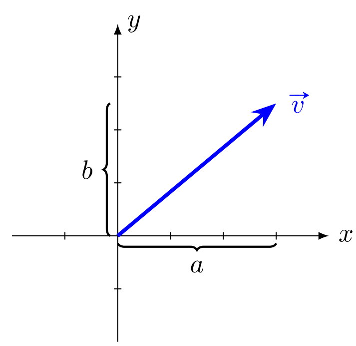


Figure 5.6: $\vec{v} = (a, b) \in \mathbb{R}^2$ such that both its components are positive.

Suppose we have $\vec{v} = (a, b) \in \mathbb{R}^2$ such that both its components are positive. We represent this situation in Figure 5.6. The length or **magnitude** of $\vec{v}$ is given as $\sqrt{a^2 + b^2}$ by the Pythagorean theorem. We can extend this definition to give the magnitude of any vector in $\mathbb{R}^n$ where n is a positive integer. We call this operation the **norm** of a vector in $\mathbb{R}^n$.

Definition 5.3. *For any vector $\vec{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$, we define the norm of the vector $\vec{w}$ to be $\|\vec{w}\| = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$.*

For $\vec{v} = (a, b) \in \mathbb{R}^2$, we have $\|\vec{v}\| = \sqrt{a^2 + b^2}$. Notice that we were careful to say that this definition of norm only relates to vectors $\in \mathbb{R}^n$. It is possible to define the norm in other ways for different spaces (even for $\mathbb{R}^n$). A vector space with a norm defined for it is, of course, called a **normed vector space**.

If we had a vector $\vec{r} \in \mathbb{R}^2$ emanating from the origin such that $\|\vec{r}\| = 1$, then the terminal point of $\vec{r}$ would be some point on the circle of radius 1 centered at the origin. Such a circle is called a *unit circle*. Any vector $\vec{u} \in \mathbb{R}^n$ such that $\|\vec{u}\| = 1$ is called a **unit vector**, so $\vec{r}$ is a unit vector $\in \mathbb{R}^2$. See Figure 5.7 for examples of some possible positions for unit vectors.

The circumference c of a circle is given by the equation $c = 2\pi r$ where $\pi = 3.1415926\ldots$ and r is the radius of the circle. For a unit circle, the

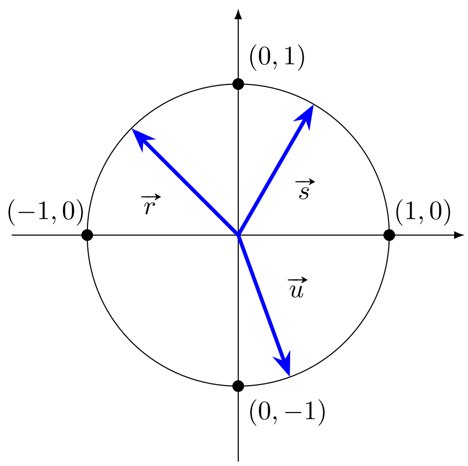


Figure 5.7: Unit Vectors in a Unit Circle

circumference is exactly equal to 2π. We define the angle that a unit vector on a plane makes with the x-axis as a proportion of the circumference of the unit circle taken from the point $(1, 0)$ in a counter-clockwise direction. Figure 5.8 gives examples of different measurements of angles. We say that these angles are given in **radians**. If the measurement is taken in a clockwise direction, then the angle is given as a negative quantity.

Sometimes angles are given in degrees, where $180° = \pi$ radians. In this text we will almost always use radian measurement.

Notice that an angle of 2π radians is the same as an angle of 0 radians or -2π radians. For our purposes we will always rewrite an angle θ as the remainder after dividing by 2π in a similar manner to how we calculate the residue of an integer modulus a prime. This means that $-2\pi \leq \theta \leq 2\pi$ for any angle θ. We use the Greek letters θ (theta), ϕ (phi) and ρ (rho) to stand for angles.[2] This is a common convention.

The angle of a unit vector in $\mathbb{R}^2$ is related to ratios of its components and its norm. These relations are so important that they are given special names and are listed below. Suppose we have $\vec{r} \in \mathbb{R}^2$ such that $\|\vec{r}\| = 1$ and $\vec{r} = (x, y)$ which makes an angle θ with the x-axis. We define:

$$\sin\theta = \frac{y}{\|\vec{r}\|} = y \qquad \cos\theta = \frac{x}{\|\vec{r}\|} = x \qquad \tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{y}{x}$$

The spoken form of these ratios are usually *sine theta*, *cosine theta* and *tan theta* (the last ratio is simply a shortened form for *tangent theta*). When one works with a coordinate system (an x and y axis, for instance), then one is doing **analytic geometry**. **Trigonometry** involves the use of the sin, cos and tan ratios.

[2]pronounced most unGreekly as THAY-TA, FIE and ROE, respectively.

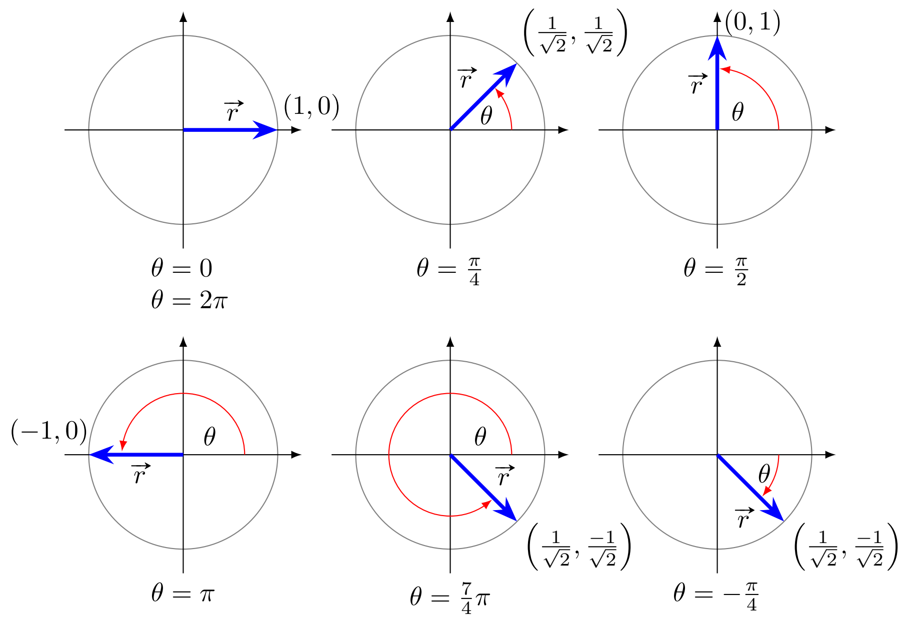


Figure 5.8: Some Radian Measurements

Suppose we have $\vec{u} = (a, b) \in \mathbb{R}^2$. Since $\|\vec{u}\|$ is a scalar, then any vector $\vec{u}$ can be written as

$$\|\vec{u}\| \frac{1}{\|\vec{u}\|} \vec{u} \quad \text{where} \quad \frac{1}{\|\vec{u}\|} \vec{u} \text{ is a unit vector.}$$

The angle this unit vector makes is taken to be the angle of the vector $\vec{u}$. The trigonometric ratios hold and are:

$$\sin\theta = \frac{b}{\|\vec{u}\|} \qquad \cos\theta = \frac{a}{\|\vec{u}\|} \qquad \tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{b}{a}$$

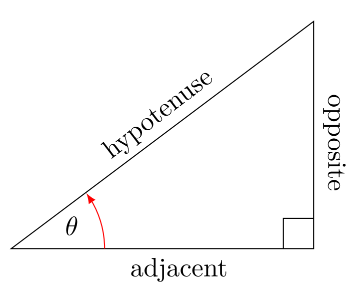


Figure 5.9: Traditional Names of the Sides of a Right-angle Triangle

Traditionally these ratios are defined using the lengths of the sides of a right-angle triangle, a triangle with one angle equal to $\frac{\pi}{2}$ or 90° (see Figure 5.9). The ratios are then defined using these lengths:

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}} \qquad \cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} \qquad \tan\theta = \frac{\text{opposite}}{\text{adjacent}}$$

Problems 5.1

1. If $\mathcal{V}$ is a vector space over $\mathbb{F}$, prove that
   (a) $c\vec{0} = \vec{0}$ for $\vec{0} \in \mathcal{V}$ and $c \in \mathbb{F}$.
   (b) $0\vec{a} = \vec{0}$ for $\vec{a} \in \mathcal{V}$, $\vec{0} \in \mathcal{V}$ and $0 \in \mathbb{F}$.
   (c) $(-1)\vec{a} = -\vec{a}$ for any $\vec{a} \in \mathcal{V}$.
2. Show that any field is a vector space over itself using the definitions of field and vector space and using the scalars of the field as vectors of the vector space.
3. Show that $\mathbb{Z}_p$, where p is a prime, is a vector space over itself.
4. Show that for $\vec{u} = (a, b) \in \mathbb{R}^2$
$$\frac{1}{\|\vec{u}\|}\vec{u} \quad \text{is a unit vector.}$$
5. In Figure 5.4, geometric examples of vector addition and subtraction in $\mathbb{R}^2$ were shown. The dashed lines formed parallelograms with the two vectors that were added or subtracted. Write out a *parallelogram law* for adding and subtracting vectors. (**HINT:** Consider that these are free vectors that can be moved around in space and joined tail-to-tip.)
6. For $\vec{v} \in \mathbb{R}^2$, the coordinates of its terminal point are $(\|\vec{v}\|\cos\theta, \|\vec{v}\|\sin\theta)$. Suppose there are two directed line segments emanating from $(1, 1)$ and having terminal points at $(2, 1)$ and $(\frac{1+\sqrt{2}}{\sqrt{2}}, \frac{1+\sqrt{2}}{\sqrt{2}})$. What is the angle θ between these two line segments? (**HINT:** First translate the line segments so that they are vectors.)
7. If $\|\vec{w} + \vec{w}\| = 2$, what are the components of $\vec{w} \in \mathbb{R}^2$?
8. If the $\vec{u} = (a, b) \in \mathbb{R}^2$ is at an angle θ from the x-axis, what are the components of a vector of the same magnitude that is $\theta + \pi$ radians from the x-axis? What about at an angle of $\theta + \frac{\pi}{2}$?
9. If $\vec{v} \in \mathbb{R}^2$ is at an angle θ from the x-axis and $\vec{w} \in \mathbb{R}^2$ is at an angle ϕ from the x-axis, the angle *between* them is given as $\rho = |\theta - \phi|$ or $\rho = 2\pi - |\theta - \phi|$, whichever is smaller. In radians, what is the angle between $\vec{v} = (1, 0)$ and $\vec{w} = (0, 1)$? If $\vec{v} = (\frac{5}{\sqrt{2}}, 0)$ and $\vec{w} = (0, \frac{1}{\sqrt{2}})$? What values can the angle between any two vectors have?

10. Let $\mathbb{F}^{\mathbf{m}\times\mathbf{n}}$ be all $m \times n$ matrices with scalars from a field $\mathbb{F}$. In other words, $M \in \mathbb{F}^{m\times n}$ means $M_{m\times n}$ (M is an $m \times n$ matrix with scalars from the field $\mathbb{F}$). Go through the definition and prove that $\mathbb{F}^{m\times n}$ is a vector space. Use matrix addition for vector addition and the product of a scalar and a matrix for the scalar product for vectors.

11. Prove that $\mathbb{R}^2$ over $\mathbb{R}$ is a vector space. Is $\mathbb{R}$ over $\mathbb{R}$ also a vector space?

12. Recall that a polynomial is any equation of the form

$$p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n = \sum_{i=0}^{n} a_ix^i$$

where $n \in \mathbb{Z} \geq 0$ and $a_i \in \mathbb{F}$. (Remember that $x^0 = 1$ for any x.) Let $\mathbb{P}$ be all polynomials $p(x)$ such that $p(x) = 0$. Prove that $\mathbb{P}$ is a vector space.

13. A vector norm is generally defined as a rule that assigns a non-negative scalar value to any vector in a vector space subject to the following conditions:

$\|\vec{x}\| > 0$ when $\vec{x} \neq \vec{0}$ and $\|\vec{x}\| = 0$ when $\vec{x} = \vec{0}$

$\|k\vec{x}\| = k\|\vec{x}\|$ for any $k \in \mathbb{F}$

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

Create some other vector norms. What might be the advantages of other norms than the one based on the Pythagorean theorem?

14. Show that the old matrix form of free vectors, as shown on Page 97, and using the standard form of matrix addition and scalar multiplication makes a vector space over any field.

15. Is there any field $\mathbb{F}$ that is *not* a vector space over itself?.

Exercises 5.1

1. What is $90°$ in radians? What is $135°$?

2. What are the magnitudes of the vectors $\vec{u} = (1, 1)$, $\vec{v} = (3, 4)$ and $\vec{w} = (-4, 3)$?

3. What are the components of $\vec{v}$ if $\|\vec{v}\| = 5$ and $\vec{v} \in \mathbb{R}^2$ is at an angle of $-\frac{\pi}{2}$ radians from the x-axis?

4. What are the components of $\vec{u}$ if $\|\vec{u}\| = 5$ and $\vec{u} \in \mathbb{R}^2$ is at an angle of $\frac{7}{4}\pi$ radians from the x-axis?

5. If $\vec{u} = (3, 4)$ and $\vec{v} = (4, 3)$, find $\vec{u} + \vec{v}$ and $\vec{u} - \vec{v}$.

6. If $\|\vec{u}\| = 3$, $\vec{u} \in \mathbb{R}^2$ is at an angle of π radians from the x-axis and $\vec{v} = (1, -1)$, find $\vec{u} + \vec{v}$ and $\vec{u} - \vec{v}$.

5.2 Inner Products: Standard Dot Product

Often we are interested in finding, on a plane, the minimum distance from a point to a line. Let us look once more at the theorem of Pythagoras and see how it can used to define the shortest distance from a point to a line.

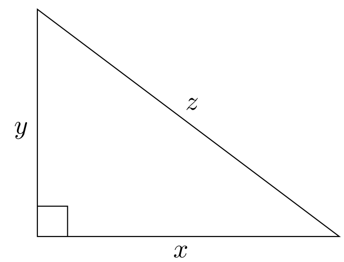


Figure 5.10: $x^2 + y^2 = z^2$

If we have a right-angle triangle such as in Figure 5.10 , the Pythagorean theorem gives us the equation $x^2 + y^2 = z^2$. Notice that z is *always* greater than or equal to y. (If $x = 0$, then $y = z$.) This is true because we are dealing with Euclidean geometry.

If we had a line that passes through two points a and b, and a point c not on the line, we can draw Figure 5.11. Notice that the dotted line segments can represent any line segment from c to the line through a and b. It should be obvious from our discussion that the line segment from c to d that is at a right angle to the line through a and b *must be the shortest line segment.* We must

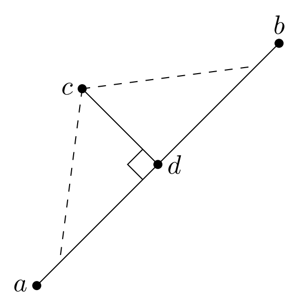


Figure 5.11: Shortest Line Segment

now get a little more complicated to show how matrix multiplication may be used to determine if a line segment is at a right angle to a line. Suppose we have a right-angle triangle composed of the points a, b and c as in Figure 5.12. We place b on the origin. Then the coordinates of the points are $a = (x_1, y_1)$, $b = (0, 0)$ and $c = (x_2, y_2)$. The labeled distances and the theorem of Pythagoras gives us the following equations:

$$x_1^2 + y_1^2 = z_1^2$$

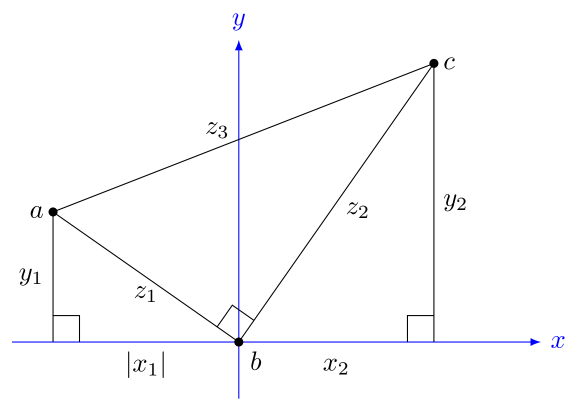


Figure 5.12: Right-Angle Triangles

$$x_2^2 + y_2^2 = z_2^2$$

Since x_1 is negative, we also have

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = z_3^2$$

Since the triangle formed by a,b and c is right-angle, we can also write:

$$z_1^2 + z_2^2 = z_3^2$$

This allows us to create the following equation:

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

which we expand to:

$$x_1^2 + y_1^2 + x_2^2 + y_2^2 = (x_2^2 - 2x_2x_1 + x_1^2) + (y_2^2 - 2y_2y_1 + y_1^2)$$

After canceling like terms we get:

$$0 = -2x_2x_1 - 2y_2y_1$$

Dividing by -2 and rearranging then gives us:

$$0 = x_1x_2 + y_1y_2$$

or, in matrix notation:

$$\begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = 0$$

This is only true when the triangle is a right-angle triangle, so we now have a way of calculating when two line segments which connect at a point are at 90° ($\frac{\pi}{2}$ radians) to each other.

Notice that we defined the coordinates of a as (x_1, y_1), a row matrix, but b is given as a column matrix. This is necessary for the calculations to work out properly. We write the coordinates of b as $(x_2, y_2)^T$ where the ^{T} means the **transpose** of a matrix. This is a convenience which we will need.

Definition 5.4. *The transpose of an $m \times n$ matrix A is an $n \times m$ matrix B such that $a_{i,j} = b_{j,i}$. Furthermore, B is written as A^T (or A is written as B^T).*

Now we can use this definition to define a type of **inner product** which is sometimes called the **standard dot product** or the **scalar product**. This type of inner product has many applications when dealing with vectors in Euclidean space.

Definition 5.5. *Suppose we are given two vectors $\in \mathbb{R}^n$, $\overrightarrow{v}$ and $\overrightarrow{w}$. We can write these vectors as equivalent column matrices $V_{n\times 1}$ and $W_{n\times 1}$. Then the dot product is the scalar $p \in \mathbb{R}$ given by the formula*

$$\overrightarrow{v} \cdot \overrightarrow{w} = V^T W = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n = p$$

We can see that $\|\overrightarrow{u}\| = \sqrt{\overrightarrow{u} \cdot \overrightarrow{u}}$ for $\overrightarrow{u} \in \mathbb{R}^n$. Furthermore, if $\overrightarrow{v} \cdot \overrightarrow{w} = 0$, we say that $\overrightarrow{v}$ and $\overrightarrow{w}$ are **orthogonal**. Also, if $\|\overrightarrow{v}\| = \|\overrightarrow{w}\| = 1$, then $\overrightarrow{v}$ and $\overrightarrow{w}$ are **orthonormal**. These definitions hold even if the norm is specified in some other way.

(**NOTE:** Unfortunately one must become accustomed, in mathematics, to words being used differently in different contexts, and this happens quite often. '*Normal*' is perhaps the most common example of a word which means different things in different contexts; the normal to a plane (or surface) in 3 dimensions is a vector perpendicular to the plane, but a normalised vector is a vector of length 1. Orthonormal vectors are perpendicular to each other but, contrary to what one might expect, that meaning comes from the prefix '*ortho-*' while the '*normal*' suffix means '*of unit length*.')

We extend the ideas of orthogonal and orthonormal to any list of vectors in $\mathbb{R}^n$. Suppose we had a list of n vectors $\overrightarrow{v_1}, \overrightarrow{v_2}, \ldots, \overrightarrow{v_n}$, and if $\overrightarrow{v_i} \cdot \overrightarrow{v_j} = 0$ for $i \neq j$, then the list of vectors is orthogonal. If we further stipulate that $\|\overrightarrow{v_i}\| = 1$, then this collection of vectors is orthonormal.

For vectors $\in \mathbb{R}^2$, two vectors are orthogonal if the angle *between* them is $\frac{\pi}{2}$ radians and they are orthonormal if they are unit vectors as well.

We are almost ready to show how to find the shortest line segment from a point to a line on the plane. First, though, we must define the **vector equation** for a line and the *projection* of a vector onto a line.

Suppose we have a line L passing through the origin. The line L consists of all points (x, y) such that $y = mx$. We can take any point on the line that is not the origin and make a vector $\overrightarrow{u}$ using this point as the vector's terminal point. Therefore any point on L can be given by scalar multiplication of the vector $\overrightarrow{u}$ by some scalar $t \in \mathbb{R}$. In other words, $L = (x, y) = t\overrightarrow{u}$ for $t \in \mathbb{R}$. See Figure 5.13.

If a line L does *not* go through the origin, we have that L is all points (x, y) such that $y = mx + b$. Here we must take two different points from L such as $u = (x_1, y_1)$ and $w = (x_2, y_2)$ and create a directed line from w to u. Then we translate this directed line segment to the vector $\overrightarrow{v} = (x_1 - x_2, y_1 - y_2)$. The points u and w can be thought of as terminal points of vectors $\overrightarrow{u}$ and $\overrightarrow{w}$, respectively. We choose one of these vectors and write that the line $L = (x, y) = \overrightarrow{w} + t\overrightarrow{v}$

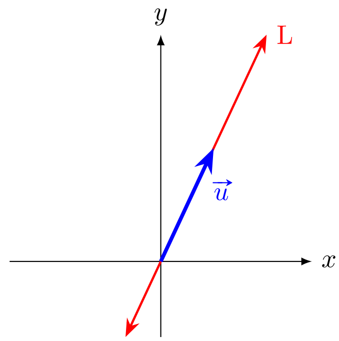


Figure 5.13: A Line through the Origin

where $t \in \mathbb{R}$. Points on L are defined uniquely by different values of t, which is called the parameter. We used the fact that any line segment which can be translated to a vector is equivalent or *parallel* to that vector. It points in the same direction. Figure 5.14 illustrates this.

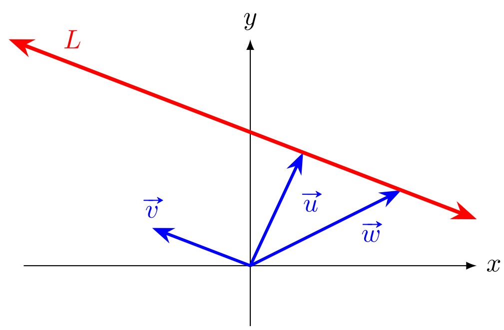


Figure 5.14: A Line not through the Origin

Notice that the vector equation for the line L is *not* unique. It depends on which points on the line we choose. Also notice that $\vec{v} = \vec{u} - \vec{w}$. This allows us to give the general definition for the vector equation of a line.

Definition 5.6. *For a line L composed of points from $\mathbb{R}^n$ and two vectors $\vec{u}, \vec{w} \in \mathbb{R}^n$ with terminal points on L, a vector equation of the line can be given as $L = \vec{w} + t(\vec{u} - \vec{w})$ where $t \in \mathbb{R}$. Another way to write the equation would be $L = \vec{u} + s(\vec{w} - \vec{u})$ where $s \in \mathbb{R}$.*

A natural question is how does one get a vector that is orthogonal to any given vector? The following theorem shows us how to do this.

Theorem 5.1. *Given a vector $\vec{w} \in \mathbb{R}^n$ such that $\|\vec{w}\| \neq 0$, then for any other vector $\vec{u} \in \mathbb{R}^n$, the vector*

$$\vec{v} = \vec{u} - \frac{\vec{u} \cdot \vec{w}}{\|\vec{w}\|^2}\vec{w}$$

is orthogonal to $\vec{w}$.

Proof.

$$\vec{v}\cdot\vec{w} = \left[\vec{u} - \frac{(\vec{u}\cdot\vec{w})\vec{w}}{\|\vec{w}\|^2}\right]\cdot\vec{w} = \vec{u}\cdot\vec{w} - \frac{(\vec{u}\cdot\vec{w})(\vec{w}\cdot\vec{w})}{\|\vec{w}\|^2}$$

$$= \vec{u}\cdot\vec{w} - \frac{(\vec{u}\cdot\vec{w})\|\vec{w}\|^2}{\|\vec{w}\|^2} = \vec{u}\cdot\vec{w} - \vec{u}\cdot\vec{w} = 0$$

□

Using this theorem, we can now define the **projection** of one vector onto another.

Definition 5.7. *Given vectors $\vec{u}$ and $\vec{w} \in \mathbb{R}^n$ with magnitudes not equal to zero, we say that the projection of $\vec{u}$ onto $\vec{w}$ is a vector $\in \mathbb{R}^n$ called $proj_{\vec{w}}\vec{u}$ defined by the formula*

$$proj_{\vec{w}}\vec{u} = \frac{\vec{u}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w}$$

The projection of a vector $\vec{u}$ onto $\vec{w}$ gives another vector which is a scalar multiple of $\vec{w}$. If $\vec{u}$ is not a scalar multiple of $\vec{w}$, then a line from the terminal point of the projection to the terminal point of $\vec{u}$ is at a right angle to the projection. If $\vec{u}$ *is* a scalar multiple of $\vec{w}$, then their terminal points and the origin are on a straight line. The reader should try a few examples to see what vector is returned by this formula in such cases.

We now have all the machinery necessary to solve our original problem, how to find the point closest to a line. Of course, the ideas we developed are useful for much more than this simple task and we will extend them to deal with more than just points and lines on a plane.

So, given a line $L = \vec{w} + t(\vec{u} - \vec{w})$ and a point $p = (x_p, y_p)$ not on the line, we can translate the line so that it passes through the origin by subtracting $\vec{w}$ from it. The vector $(\vec{u} - \vec{w})$ points in the same direction as this line. We can also create a vector $\vec{p}$ which has a terminal point at (x_p, y_p). Then the vector $(\vec{p} - \vec{w})$ has a terminal point that is also translated the same distance. We project this vector $(\vec{p} - \vec{w})$ onto the vector $(\vec{u} - \vec{w})$. The projected vector is $\text{proj}_{(\vec{u}-\vec{w})}(\vec{p} - \vec{w})$. Translate this vector onto the line L by adding the vector $\vec{w}$ and the result is a vector with terminal point on the line L that is the shortest distance from the point p. This is illustrated in Figure 5.15. The unlabeled vector in the bottom right graph in Figure 5.15 is the vector $\text{proj}_{(\vec{u}-\vec{w})}(\vec{p} - \vec{w})$. We can see that $\vec{q} = \text{proj}_{(\vec{u}-\vec{w})}(\vec{p} - \vec{w}) + \vec{w}$.

As was mentioned previously, a physicist usually views all directed line segments as vectors so that the addition of vectors is accomplished by imagining one vector being placed so that its emanating point is the terminal point of the other vector and the sum is the directed line segment from the unconnected emanating point to the unconnected terminal point. As well, projections may

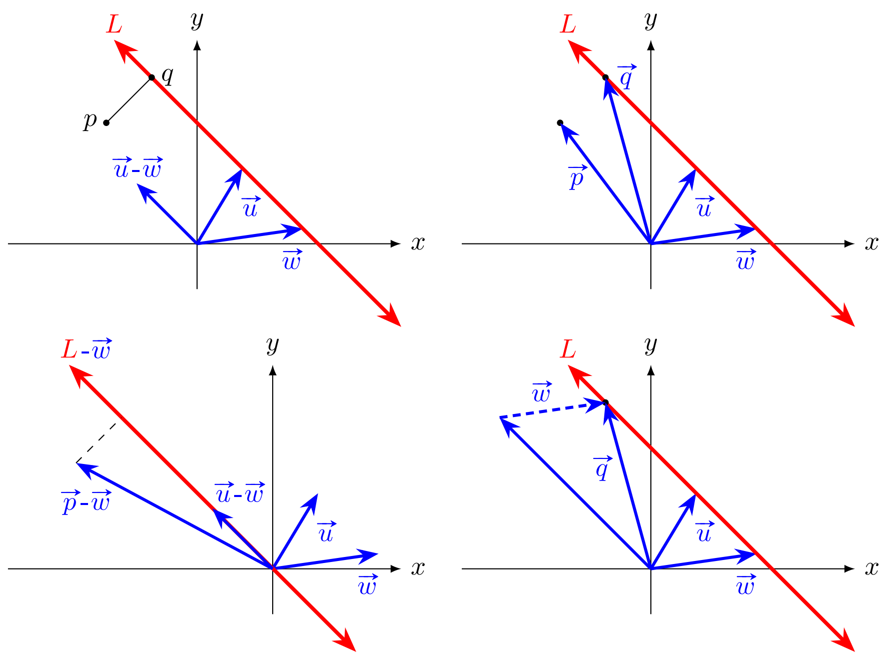


Figure 5.15: A Line L and a Point p not on L

occur on line segments that are not vectors emanating from the origin. Viewing things in this way can give a better feeling for vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ but the amount of calculation required to get an answer remains the same. In order to illustrate a calculation, we let $\vec{u} = (1, 3)$ and $\vec{w} = (4, 2)$. Then the line L is given by $L = \vec{w} + t(\vec{u} - \vec{w})$ or $L = (4, 2) + t(-3, 1)$. We also let $\vec{p} = (-2, 4)$ and we want to find the point q which is on L that is closest to the terminal point of $\vec{p}$.

Our formula gives us

$$\begin{aligned}
\vec{q} &= \text{proj}_{(\vec{u}-\vec{w})}(\vec{p}-\vec{w}) + \vec{w} \\
&= \text{proj}_{(-3,1)}(-6,2) + (4,2) \\
&= \frac{(\vec{p}-\vec{w}) \cdot (\vec{u}-\vec{w})}{\|(\vec{u}-\vec{w})\|^2}(\vec{u}-\vec{w}) + \vec{w} \\
&= \frac{(-6,2) \cdot (-3,1)}{\|(-3,1)\|^2}(-3,1) + (4,2) \\
&= \frac{(-6,2) \cdot (-3,1)}{(-3,1) \cdot (-3,1)}(-3,1) + (4,2) \\
&= \frac{(-6) * (-3) + (2) * (1)}{(-3) * (-3) + (1) * (1)}(-3,1) + (4,2) \\
&= \frac{20}{10}(-3,1) + (4,2) \\
&= 2 * (-3,1) + (4,2) \\
&= (-2,4)
\end{aligned}$$

We see that $\vec{q} = \vec{p}$ which means that p is on the line L. We can check this by converting our equation for L into point-intercept form. We start with the fact that $L = (4,2) + t(-3,1)$ is really a compact notation for two equations:

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

These equations can be solved for t:

$$\begin{aligned}
t &= \frac{4-x}{3} \\
t &= y - 2
\end{aligned}$$

This gives us the linear equation:

$$y = \frac{-1}{3}x + \frac{10}{3}$$

For $x = -2$ and $y = 4$, it is easy to show that this equation holds.

If the point $p = (-1,7)$, then $\text{proj}_{(\vec{u}-\vec{w})}(\vec{p}-\vec{w}) + \vec{w} = \text{proj}_{(-3,1)}(-5,5) + (4,2) = (-2,4)$. This means that $(\vec{p} - \vec{q})$, the vector equivalent to the directed line segment from q to p, is orthogonal to the vector $(\vec{u} - \vec{w})$. It is easy to verify this using the standard dot product.

In this text we stated that a vector $\vec{a}$ in $\mathbb{R}^n$ can be represented by a matrix $A_{1\times n}$ and that the dot product of two vectors in $\mathbb{R}^n$, $\vec{a} \cdot \vec{b}$, is equivalent to the matrix multiplication AB^T, which returns a scalar $\in \mathbb{R}$. We have also shown a geometric interpretation for points and lines on a plane.

It is possible to show other geometric interpretations of vectors and vector spaces. We will do so in the following sections not only for their utility, but also as an aid to understanding some of the beautiful theorems of linear algebra which may appear very abstract at first.

Problems 5.2

1. A vector space with an inner product that associates each pair of vectors in that space with a non-negative scalar quantity[3] is called an **inner product space**. Could one define an inner product for vector spaces over other fields besides the reals?

2. For vectors in $\mathbb{R}^n$, prove that $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{v}$. (**HINT:** Replace the vectors with matrices and use the summation notation of matrix multiplication.) This proves that the dot product is distributive with respect to addition.

3. Suppose $\vec{v}, \vec{w} \in \mathbb{R}^2$ and θ is the angle between them. Show that $\vec{v} \cdot \vec{w} = \|\vec{v}\|\|\vec{w}\| \cos\theta$. (**HINT:** Use $\text{proj}_{\vec{w}}\vec{v}$, the length of $\vec{v}$ and the traditional definition of $\cos\theta$.)

4. Show that the **zero vector**, $\vec{0}$, which has components which are all zero, is orthogonal to any other vector. What is the relationship between $\vec{0}$ and $\mathbf{0}$?

5. A line $L = \vec{q} + r\vec{s}$ is considered parallel to a line $M = \vec{t} + u\vec{v}$ if $\vec{s} = w\vec{v}$ for some $w \in \mathbb{R}$. Show that for two non-zero vectors $\vec{a}$ and $\vec{c}$ that a line containing $\text{proj}_{\vec{c}}\vec{a}$ is parallel to a line containing $\vec{c}$.

6. For $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$, show that $\vec{u} - \text{proj}_{\vec{v}}\vec{u}$ is orthogonal to $\vec{v}$.

7. Suppose that $\vec{i}$ and $\vec{j} \in \mathbb{R}^2$ are orthonormal. Show that for any real numbers s and t, the vectors $\vec{u} = s\,\vec{i} + t\,\vec{j}$ and $\vec{u} = t\,\vec{i} - s\,\vec{j}$ are orthogonal.

8. Suppose that $\vec{i}$ and $\vec{j} \in \mathbb{R}^2$ are orthonormal. Show that any vector $\vec{u} \in \mathbb{R}^2$ can be written as $\vec{u} = s\,\vec{i} + t\,\vec{j}$ for some $s, t \in \mathbb{R}$.

9. Prove that $(A^T)^T = A$ for any matrix A.

10. If A and B are both $m \times n$ matrices, show that $(A + B)^T = A^T + B^T$.

11. Assume $A_{m\times n}$ and $B_{n\times p}$ are two matrices over $\mathbb{F}$. Using the definition of matrix multiplication, prove that $(AB)^T = B^T A^T$.

12. Show that $(aB)^T = aB^T$ for any scalar a.

13. Prove that if A is a nonsingular matrix, then so is A^T and that $(A^T)^{-1} = (A^{-1})^T$.

14. If E is an elementary matrix, show that E^T is also an elementary matrix.

[3]with some other conditions which we will not list here.

15. Let $A_{2\times 2}$ with entries in $\mathbb{R}$ be a matrix such that each column is a unit vector and the columns are orthogonal. Prove that A is invertible and that $A^{-1} = A^T$. We say that A is an **orthogonal matrix** .

Exercises 5.2

1. For vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^2$, compute the dot products. Calculate $\|\vec{v} - \vec{u}\|$.
   (a) $\vec{u} = (2,3),\ \vec{v} = (1,5)$
   (b) $\vec{u} = (0,0),\ \vec{v} = (-1,3)$
   (c) $\vec{u} = (-2,1),\ \vec{v} = (1,2)$
   (d) $\vec{u} = (0,-3),\ \vec{v} = (0,-3)$

2. For the vectors in the previous problem, find $\text{proj}_{\vec{u}}\vec{v}$. Find $\text{proj}_{\vec{v}}\vec{u}$.

3. In a plane, for a line L given by the vector equation $L = \vec{u} + t(\vec{w} - \vec{u})$ and a point p, calculate the closest point on L to p, given that
   (a) $\vec{u} = (1,1),\ \vec{w} = (2,1)$
   $p = (0,4)$
   (b) $\vec{u} = (0,0),\ \vec{w} = (1,0)$
   $p = (0,4)$
   (c) $\vec{u} = (1,0),\ \vec{w} = (0,1)$
   $p = (3,4)$
   (d) $\vec{u} = (1,-3),\ \vec{w} = (1,3)$
   $p = (-1,-4)$

4. Using the procedure for matrix multiplication that was developed for Lampp, create a similar algorithm for calculating the projection of one vector onto another.

5. On a scrap of paper, write out the definition of dot product and the definition of the projection of one matrix onto another 50 times.

5.3 The Outer Product

Just as we have directed line segments, it is possible to define *directed areas*. This may seem a strange concept initially, but it is really quite straightforward. First, let us consider what we mean, in a little more detail, when we talk about area. If we have a rectangle 2 meters wide and 3 meters in height, we say that its area is $2 * 3 = 6$ square meters. In other words, it has an area *equivalent to 6 unit squares*. Since ordinary multiplication is commutative, the order in which we multiply doesn't matter, so $2 * 3 = 3 * 2 = 6$. When we give an area as a simple real number, we have lost the information about the order in which we multiplied. Sometimes this information is valuable, as we shall see. One way to keep this information in our calculations is to use **unit bivectors**. Look at Figure 5.16 and match it to the following definition.

Definition 5.8. *A unit bivector is the area of the square created by two orthonormal vectors and a sign indicating the order in which the area was calculated. If*

$\vec{x}$ and $\vec{y}$ are two orthogonal unit vectors, then two unit bivectors may be created. They are called $\vec{x} \wedge \vec{y}$ and $\vec{y} \wedge \vec{x}$. Furthermore, $\vec{x} \wedge \vec{y} = -(\vec{y} \wedge \vec{x})$.

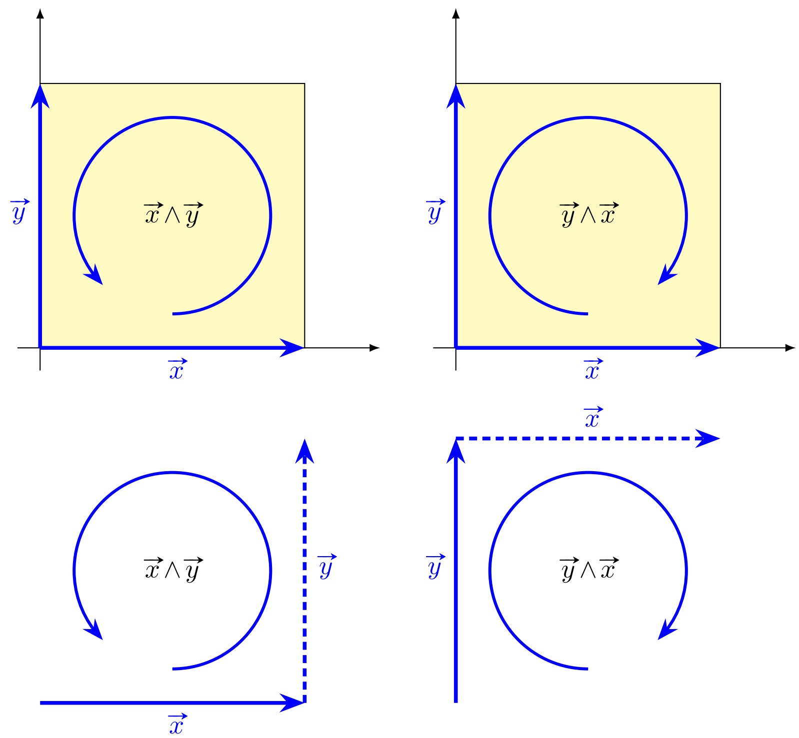


Figure 5.16: $\vec{x}\wedge\vec{y}$ and $\vec{y}\wedge\vec{x}$

The bottom two sketches in Figure 5.16 show an easy way of remembering the orientation (direction) of a unit bivector.

The symbol $\wedge$ is used to denote the **outer product**, sometimes referred to as the **wedge product**, so that $\vec{u} \wedge \vec{w}$ would be spoken as *u wedge w*.

When we have drawn lines and planes, we have made use of a coordinate system. It is not always necessary or even desirable to do so. The definition of a unit bivector refers to two orthonormal vectors, which do not have to be parallel to the x and y axes. When we refer to unit vectors, we will adopt the convention of representing them as $\vec{e}_1$, $\vec{e}_2$, etc. A unit bivector made from the two unit vectors $\vec{e}_1$ and $\vec{e}_2$ will be represented by $\vec{e}_{12}$ or $\vec{e}_{21}$ such that

$$\vec{e}_1 \wedge \vec{e}_2 = \vec{e}_{12}$$

$$\vec{e}_2 \wedge \vec{e}_1 = \vec{e}_{21}$$

It should be apparent from the previous definition that $\vec{e}_{12} = -\vec{e}_{21}$.

A **parallelogram** is a four-sided object on the plane such that the length of opposite sides are equal. This also means that opposite angles are equal.

The simplest kinds of parallelograms are squares and rectangles. The area of a rectangle is easy to calculate and we would like a formula which will give us the area of a parallelogram. We will use Figure 5.17 as a guide to develop a formula.

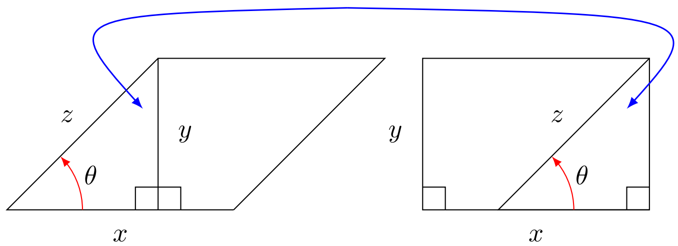


Figure 5.17: A Parallelogram and its Equivalent Rectangle

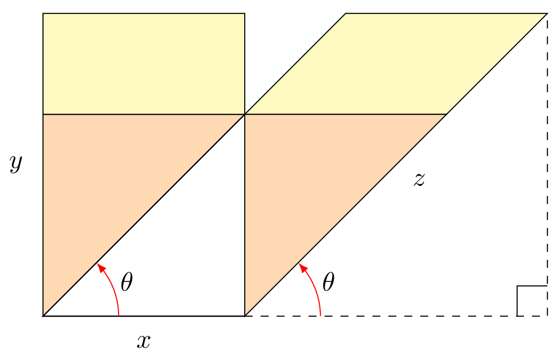


Figure 5.18: Arranging a Parallelogram into its Equivalent Rectangle

Figure 5.17 does not encompass situations such as Figure 5.18. Figure 5.18 shows how such a parallelogram can be cut up and rearranged into an equivalent rectangle. These two diagrams represent all parallelograms. The parallelogram in yellow at the top of Figure 5.18 is similar to the parallelogram given in Figure 5.17. From the definition of sin we have that $\sin\theta = y/z$ or $y = z\sin\theta$. This lets us use algebra to give the formula that

$$\text{The area of a parallelogram} = xy = xz\sin\theta$$

The area of the parallelogram we calculated is *not* a directed area. In order to use this result to talk about directed areas, we need to reformulate our geometrical exercise into the language of vectors and bivectors. If we replace the line segments in our diagrams with vectors, we have three vectors $\vec{x}$, $\vec{z}$ and $\vec{y}$. The norms of the vectors are equal to the lengths we stipulated in the geometrical diagrams so that $\|\vec{x}\| = \text{x}$, etc.

We have seen previously that scalar multiplication of a vector by a positive scalar value results in a change in the magnitude of the vector but does not affect its direction. If we have a vector in $\mathbb{R}^2$, for example, that is parallel to a unit vector $\vec{e}_1$, then this vector is $\vec{x} = \|\vec{x}\|\vec{e}_1$. The vector $\vec{x}$ is a scalar multiple of a unit vector. We can now see that the outer product of two vectors on a plane is a scalar multiple of a unit bivector. This gives us a directed area. Notice that the unit vectors used to define unit bivectors are not necessarily parallel to coordinate axes. We do need to adopt the convention that $\vec{e}_{12}$ is a unit bivector that has a counter-clockwise orientation. This means, of course, that $\vec{e}_{21}$ has a clockwise direction. Given two vectors, $\vec{a}$ and $\vec{c}$, both in $\mathbb{R}^2$, and two unit bivectors, then $\vec{a} \wedge \vec{b}$ will either equal $\|\vec{a}\|\|\vec{c}\| \sin\theta\, \vec{e}_{12}$ if the angle between $\vec{a}$ and $\vec{c}$ is measured in the same direction as $\vec{e}_{12}$ or it will equal $\|\vec{a}\|\|\vec{c}\| \sin\theta\, \vec{e}_{21}$. Remembering that $\sin 0 = 0$ shows us that $\vec{e}_{11}$ and $\vec{e}_{22}$ always have zero magnitude.

Using parallelograms, we can convince ourselves that the outer product of unit vectors in $\mathbb{R}^2$ is distributive with respect to vector addition. In other words, that

$$\vec{e}_1 \wedge (\vec{e}_1 + \vec{e}_2) = \vec{e}_1 \wedge \vec{e}_1 + \vec{e}_1 \wedge \vec{e}_2 = \vec{e}_1 \wedge \vec{e}_2 = \vec{e}_{12}$$

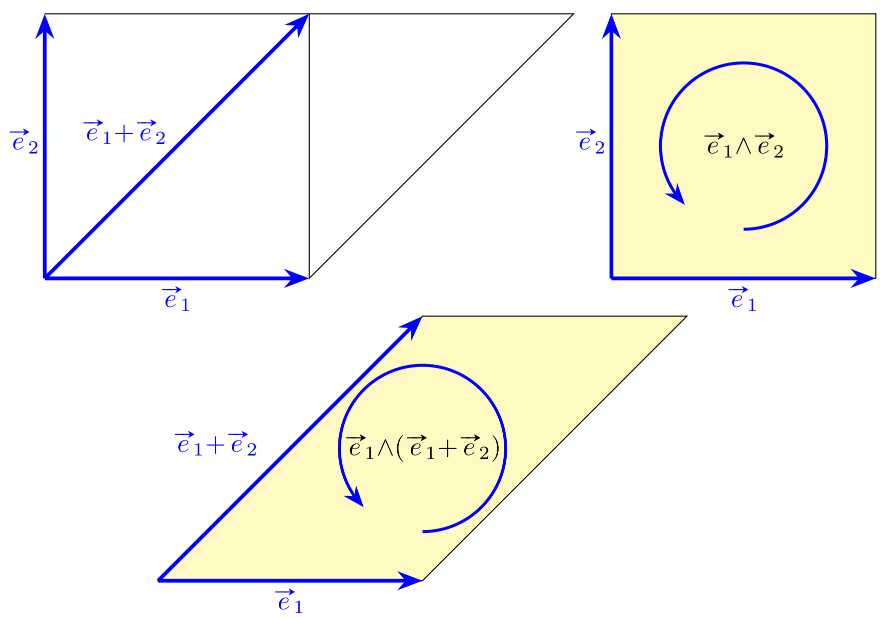


Figure 5.19: $\vec{e}_1\wedge(\vec{e}_1+\vec{e}_2) = \vec{e}_1\wedge\vec{e}_1+\vec{e}_1\wedge\vec{e}_2$

Notice that since $\vec{e}_1 \wedge \vec{e}_1$ has zero magnitude, we remove it from the final expression on the right. (We haven't really proved anything in a formal sense. We will do so later when we introduce more general definitions of the inner and outer product. Our purpose here is to show how geometric insights can lead to conjectures which we can then strengthen into theorems.)

For any two vectors $\vec{u}$ and $\vec{w}$ in $\mathbb{R}^2$, we can use one of the vectors to define a unit vector, so that $\|\vec{u}\|\vec{e}_1 = \vec{u}$, for instance. Then the second vector $\vec{w}$ can be written in terms of the angle between it and the other vector and two orthonormal vectors.

$$\vec{w} = \|\vec{w}\| \cos\theta \vec{e}_1 + \|\vec{w}\| \sin\theta \vec{e}_2$$

Figure 5.20 shows an example of two such vectors. Notice that the angle θ is measured in the same direction as $\vec{e}_{12}$. We say that $\vec{w}$ is a **linear combination** of the orthonormal vectors $\vec{e}_1$ and $\vec{e}_2$.

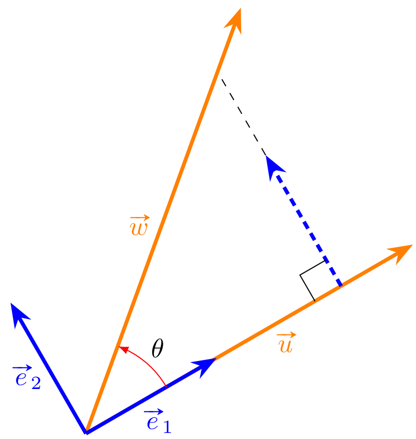


Figure 5.20: Two Vectors in Reference to Two Orthonormal Vectors

Definition 5.9. *A vector $\vec{a}$ from a vector space $\mathcal{V}$ over a field $\mathbb{F}$ is a linear combination of n vectors $\vec{c}_i$ from $\mathcal{V}$ if*

$$\vec{a} = b_1 \vec{c}_1 + b_2 \vec{c}_2 + \cdots + b_n \vec{c}_n$$

where b_i are scalars $\in \mathbb{F}$.

Letting $\vec{u}$ be a scalar multiple of a unit vector, writing $\vec{w}$ as a linear combination of two orthonormal unit vectors and using the distributive property of the outer product for unit vectors allows us to write the outer product of two vectors in $\mathbb{R}^2$ as

$$\begin{aligned}
\vec{u} \wedge \vec{w} &= \|\vec{u}\|\vec{e}_1 \wedge (\|\vec{w}\| \cos\theta \vec{e}_1 + \|\vec{w}\| \sin\theta \vec{e}_2) \\
&= \|\vec{u}\|\vec{e}_1 \wedge \|\vec{w}\| \cos\theta \vec{e}_1 + \|\vec{u}\|\vec{e}_1 \wedge \|\vec{w}\| \sin\theta \vec{e}_2 \\
&= \|\vec{u}\|\|\vec{w}\| \cos\theta \vec{e}_1 \wedge \vec{e}_1 + \|\vec{u}\|\|\vec{w}\| \sin\theta \vec{e}_1 \wedge \vec{e}_2 \\
&= \|\vec{u}\|\|\vec{w}\| \sin\theta \vec{e}_1 \wedge \vec{e}_2 \\
&= \|\vec{u}\|\|\vec{w}\| \sin\theta \vec{e}_{12}
\end{aligned}$$

We can extend the ideas of unit bivectors to unit **trivectors** and further on into unit **multivectors**. As we are interested only in developing enough ideas to motivate a basic understanding of linear algebra, we will confine ourselves to three dimensions and the trivector. The following definition makes use of the properties of the unit bivector.

Definition 5.10. *A unit trivector is the volume of the cube created by three orthonormal vectors and a sign indicating an orientation or direction of the volume. If $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ are three orthogonal unit vectors in $\mathbb{R}^3$, then two unit trivectors may be created using the outer product. They are called $\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3$, written as $\vec{e}_{123}$, and $\vec{e}_3 \wedge \vec{e}_2 \wedge \vec{e}_1$, written as $\vec{e}_{321}$. Furthermore, $\vec{e}_{123} = -\vec{e}_{321}$.*

A unit trivector is a bit of three-dimensional space with a potential twist or spin. One analogy is a screw, which can be threaded right-handed or left-handed. The direction one turns a screw to get it to screw into a surface does not change with the position of the screw.

One can think of a unit bivector as the area swept out by moving a unit vector's originating point along the another unit vector from its originating point to its terminal point. The same can be said for a unit trivector, where a unit bivector composed of two unit vectors is swept along the third unit vector to give a volume. In this way we can see geometrically that the outer product of unit vectors is associative, that

$$\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3 = (\vec{e}_1 \wedge \vec{e}_2) \wedge \vec{e}_3 = \vec{e}_1 \wedge (\vec{e}_2 \wedge \vec{e}_3)$$

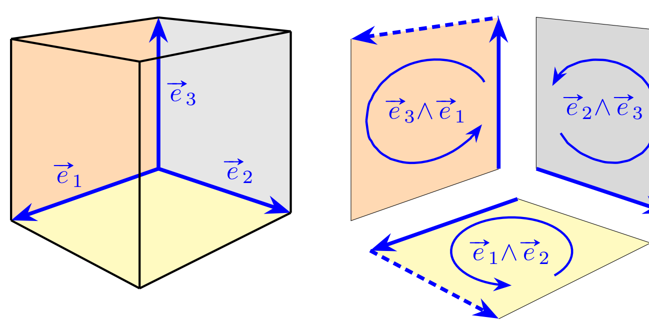


Figure 5.21: A Trivector with Bivector Components

Every time we use the outer product on orthonormal vectors, we are creating an object that is one dimension greater than the previous object. So, $\vec{e}_1$ is a directed line segment, $\vec{e}_1 \wedge \vec{e}_2$ is a directed area and $\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3$ gives a directed volume.

We will leave these geometrical constructions for now and return to them from time to time in order to illustrate general theorems in linear algebra.

Problems 5.3

1. If $\vec{u}$, $\vec{v}$ and $\vec{w}$ are vectors in $\mathbb{R}^2$ such that $\vec{u} = \vec{v} + c\vec{w}$ for $c \in \mathbb{R}$, prove that $\vec{u} \wedge \vec{w} = \vec{v} \wedge \vec{w}$.

2. Show that for any two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^2$ that $\vec{v} \wedge \vec{w} = -\vec{w} \wedge \vec{v}$.

3. Assume $\vec{e}_i$ are *orthogonal* vectors in $\mathbb{R}^3$, $\vec{u} = u_1\vec{e}_1 + u_2\vec{e}_2 + u_3\vec{e}_3$ and that $\vec{w} = w_1\vec{e}_1 + w_2\vec{e}_2 + w_3\vec{e}_3$, where u_i and $w_i \in \mathbb{R}$. Show that $\vec{u} \wedge \vec{w} = (u_1w_2 - u_2w_1)\vec{e}_1 \wedge \vec{e}_2 + (u_2w_3 - u_3w_2)\vec{e}_2 \wedge \vec{e}_3 + (u_3w_1 - u_1w_3)\vec{e}_3 \wedge \vec{e}_1$. Take for granted that the outer product for orthogonal vectors in $\mathbb{R}^3$ is distributive.

5.4 Subspaces

Definition 5.11. *Suppose we use $\mathcal{S}$ to name a collection of one or more vectors picked from a vector space $\mathcal{V}$ over $\mathbb{F}$. Using the same addition and multiplication as for vectors in $\mathcal{V}$, then $\mathcal{S}$ is called a* ***subspace*** *of $\mathcal{V}$ if $\mathcal{S}$ is also a vector space over $\mathbb{F}$.*

The manipulative properties which define a vector space are carried over from $\mathcal{V}$ to $\mathcal{S}$, so the associative, commutative and distributive properties for vectors in $\mathcal{S}$ still apply. We therefore only have to check for the existence of certain vectors in $\mathcal{S}$. Specifically, we need to check:

1. Is $\mathcal{S}$ closed under vector addition? In other words, for every $\vec{u}$, $\vec{w}$ in $\mathcal{S}$, is $\vec{u} + \vec{w}$ in $\mathcal{S}$?

2. Is $\mathcal{S}$ closed under scalar multiplication? For every $c \in \mathbb{F}$ and every $\vec{u}$ in $\mathcal{S}$, is $c\vec{u}$ in $\mathcal{S}$?

If the answer to any one of these questions is no, then $\mathcal{S}$ is not a subspace and we cannot treat it as a vector space over $\mathbb{F}$.

Notice that the second condition assures us that $\vec{0}$ is in $\mathcal{S}$. It also assures us that if $\vec{w}$ is in $\mathcal{S}$, then $-1\vec{w} = -\vec{w}$ is in $\mathcal{S}$.

We can use a theorem to give us a test to see if a collection of vectors from a vector space is really a subspace.

Theorem 5.2. *Let $\mathcal{V}$ be a vector space over $\mathbb{F}$ and $\mathcal{S}$ is a collection of one or more vectors from $\mathcal{V}$, then $\mathcal{S}$ is a subspace of $\mathcal{V}$ if and only if any linear combination of any two vectors in $\mathcal{S}$ is also in $\mathcal{S}$. In other words, if a,$b \in \mathbb{F}$ and $\vec{u}$, $\vec{w} \in \mathcal{S}$, then $a\vec{u} + b\vec{w}$ must also be in $\mathcal{S}$.*

Proof. If $\mathcal{S}$ is a vector space, then it certainly satisfies this property, since we have $(a+b)(\vec{u}+\vec{w})$ must also be in $\mathcal{S}$. Assuming that this closure property holds, then $\mathcal{S}$ must be a vector space since it satisfies the two tests given. If $a = b = 1$, then the first test is satisfied. If $a = 0$, then the second test is satisfied. □

Any vector space is a subspace of itself. There is certainly no need to check its properties.

The simplest example of a subspace is just the vector $\vec{0}$. The properties of a vector space assures us that this subspace is a subspace of *every* subspace. This subspace is called the **zero subspace**.

Any subspace of a vector space other than the vector space itself or the zero subspace is called a **proper subspace** .

Suppose we had several subspaces of a vector space $\mathcal{V}$. Combining all vectors that are common to all subspaces is called taking the **intersection** of the subspaces. We use the symbol $\cap$ to refer to the intersection of two subspaces so that $\mathcal{S} = \mathcal{S}_1 \cap \mathcal{S}_2$ means that $\mathcal{S}$ contains all vectors common to the two subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$. Another way of saying this: if $\mathcal{S} = \mathcal{S}_1 \cap \mathcal{S}_2$, $\forall s \in \mathcal{S}$, $s \in \mathcal{S}_1$ and $s \in \mathcal{S}_2$. (If $\mathcal{S}$ is equal to the intersection of $\mathcal{S}_1$ and $\mathcal{S}_2$, **for all** s in $\mathcal{S}$, s is in $\mathcal{S}_1$ and s is in $\mathcal{S}_2$.)

Theorem 5.3. *If $\mathcal{V}$ is a vector space over $\mathbb{F}$ and $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \ldots, \mathcal{S}_n$ are subspaces of $\mathcal{V}$, then so is*

$$\mathcal{S} = \cap_{i=1}^{n} \mathcal{S}_i = \mathcal{S}_1 \cap \mathcal{S}_2 \cap \cdots \cap \mathcal{S}_n$$

Proof. If $\vec{u}$ and $\vec{w}$ are in $\mathcal{S}$, then they are also members of each $\mathcal{S}_i$, for $i = 1, 2, \ldots, n$. Since each $\mathcal{S}_i$ is a subspace, then for $a, c \in \mathbb{F}$, $a\vec{u} + c\vec{w}$ are also in $\mathcal{S}_i$, for $i = 1, 2, \ldots, n$. Therefore $a\vec{u} + c\vec{w}$ is in $\mathcal{S}$. By the previous theorem, this assures us that $\mathcal{S}$ is a subspace. $\square$

Recall that we can write a system of equations in the form $AX = K$ where A is an $m \times n$ matrix and K is an $m \times 1$ matrix. The **solution space** of this system is all $n \times 1$ matrices X which satisfies this equation. If $K = \mathbf{0}$, then this matrix equation is called a **homogeneous system**. Otherwise it is referred to as nonhomogeneous.

The solution space of the homogeneous system $AX = \mathbf{0}$ is often called the **null space** of A and we refer to it as $\mathcal{NS}(A)$. If $\mathcal{NS}(A)$ exists, then it forms a subspace. This subspace is the most important subspace we will encounter. To prove that it is a subspace we will use the result of a previous problem where we proved that $\mathbb{F}^{m \times n}$, all $m \times n$ matrices over a field $\mathbb{F}$, is a vector space. (If you didn't do the problem, feel guilty for a moment, then go back and do it now.) Obviously X is a vector in this vector space. Suppose we have $AX_1 = \mathbf{0}$ and $AX_2 = \mathbf{0}$ are any solutions to the homogeneous system. For any $a, b \in \mathbb{F}$, $aAX_1 = A(aX_1) = \mathbf{0}$ and, similarly, $A(bX_2) = \mathbf{0}$. Then $A(aX_1) + A(bX_2) = \mathbf{0}$ and $A(aX_1 + bX_2) = \mathbf{0}$. (Again, showing that matrix multiplication is associative was an assigned problem.) Therefore $(aX_1 + bX_2)$ is a vector in the solution space, so we have shown $\mathcal{NS}(A)$ is a subspace. $\mathcal{NS}(A)$ is sometimes also referred to as the **kernel** of A.

There exists a very nice relationship between homogeneous and nonhomogeneous systems. Suppose we have a nonhomogeneous system $AX = B$ and two solutions, X_1 and X_2. By the distributive law of matrix multiplication, we have $A(X_1 - X_2) = AX_1 - AX_2 = B - B = \mathbf{0}$. This means that $Y = X_1 - X_2$ is a solution of the homogeneous system $AX = \mathbf{0}$. Looking at this another way, if X_1 is a particular solution of the nonhomogeneous system $AX = B$ and X_2 is any other solution of $AX = B$, then there exists a Y, a solution of the homogeneous system $AX = \mathbf{0}$, such that $X_2 = X_1 + Y$. This means that we can get all solutions to $AX = B$ by finding one particular solution and then finding *all* solutions to $AX = \mathbf{0}$.

One geometric example of this would be the intersection of two planes which do not go through the origin. We would first use the same translation to move each of the planes so that they go through the origin. Then we would find the line of intersection, then translate this line back.

It is important to remember that a vector space is not simply a collection of vectors, but that it must also have two binary operations defined. These operations are 'inherited' by a subspace. For example, the vector space $\mathbb{Z}_2$ is simply composed of the two integers 0 and 1. Yet $\mathbb{Z}_2$ is not a subspace of $\mathbb{R}$, even though $\mathbb{R}$ contains these two integers, since addition is defined differently for these two vector spaces.

Problems 5.4

1. If $AX = K$ and $K \neq 0$, what can we say about $\mathcal{NS}(A)$?

2. Show that the solution space of the nonhomogeneous system $AX = K$, where K is not the zero subspace, is not a subspace.

3. Suppose we have $\mathcal{S}$ is all real points (x, y) such that $y = mx$ for one particular $m \in \mathbb{R}$. This is a line in the Euclidean plane that passes through the origin. Show that $\mathcal{S}$ is a subspace of $\mathbb{R}^2$. Is $\mathcal{S}$ a proper subspace of $\mathbb{R}^2$? (**HINT**: Any point (x, y) on the plane can be represented as a vector in $\mathbb{R}^2$ with its terminal point at (x, y)).

4. Show that any line in the Euclidean plane that does not go through the origin cannot be a subspace of $\mathbb{R}^2$. (**HINT**: Consider the existence of $\vec{0}$.)

5. Is $\mathbb{Z}_2$ a subspace of $\mathbb{Z}_3$?

6. Let $\mathcal{S}$ be all the points (x, y, z) such that $x = at$, $y = bt$ and $z = ct$ where $a, b, c, t \in \mathbb{R}$. Then $\mathcal{S}$ consists of the vectors in $\mathbb{R}^3$ lying on a straight line passing through the origin. Let $\vec{u} = (at_1, bt_1, ct_1) \in \mathcal{S}$ and $\vec{w} = (at_2, bt_2, ct_2) \in \mathcal{S}$. Use these two vectors to show that $\mathcal{S}$ is a subspace of $\mathbb{R}^3$.

7. Show that the set of points in $\mathbb{R}^3$ lying on a plane passing through the origin is a vector space. (**HINT**: We want to show that the points (x, y, z) such that $ax + by + cz = 0$ are a subspace of $\mathbb{R}^3$.)

8. A **diagonal matrix** is an $n \times n$ matrix with all of its off-diagonal entries equal to zero. Prove that $\mathcal{D}_n$, all $n \times n$ diagonal matrices over a field $\mathbb{F}$, is a subspace of $\mathbb{F}^{n \times n}$.

9. An $n \times n$ matrix U is called **upper triangular** if $u_{i,j} = 0$ whenever $i > j$. Show that $\mathcal{U}_n$, all such matrices U over a field $\mathbb{F}$, is a subspace of $\mathbb{F}^{n \times n}$. If $L = U^T$, then L is a **lower triangular** matrix. Show that $\mathcal{L}_n$, all such matrices L over a field $\mathbb{F}$, is also a subspace of $\mathbb{F}^{n \times n}$.

10. Prove that $\mathbb{R}$ has no proper subspace.

11. Is $\mathbb{R}$ a proper subspace of $\mathbb{R}^2$?

12. A polynomial is said to have degree n if n is the largest exponent in the polynomial. Let $\mathbb{P}^n$ be all polynomials of degree less than or equal to to n. Prove that $\mathbb{P}^n$ is a subspace of $\mathbb{P}$, the vector space of all polynomials.

13. Prove that $\mathbb{R}$ is a proper subspace of $\mathbb{C}$.

14. Is $\mathbb{R}^2$ a proper subspace of $\mathbb{C}$? Of $\mathbb{C}^2$?

15. Show that $\mathbb{Z}_p$, for p a prime, does not have a proper subspace.

This section may appear a bit abstract and far removed from many of the ideas of matrix manipulation which we have covered. This is not true at all and we will discuss a situation which will begin to tie together many of the concepts we have already seen.

We will examine two particular subspaces of $\mathbb{R}^3$ and show that their intersection is also a subspace. In $\mathbb{R}^3$ let $\mathcal{S}_1$ be all points (x, y, z) that satisfies the equation $x + \frac{3}{2}y + z = 0$. We have seen that this equation determines a plane in $\mathbb{R}^3$ that passes through the origin. This means $\mathcal{S}_1$ is a subspace of $\mathbb{R}^3$ (from the result of a problem in this section). We also let $\mathcal{S}_2$ be a subspace of $\mathbb{R}^3$ defined as all points (x, y, z) that satisfies the equation $2x+2y+z = 0$. Then the intersection of these two subspaces, $\mathcal{S}_1 \cap \mathcal{S}_2$, is all points (x, y, z) which satisfies both equations. We can write this as a homogeneous system of equations.

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

We can represent this system of equations by the matrix equation $AX = K$ such that

$$A = \begin{pmatrix} 1 & \frac{3}{2} & 1 \\ 2 & 2 & 1 \end{pmatrix} \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad K = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

We use an augmented matrix containing A and K to create the associated matrix for this system.

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

This matrix can be row reduced (using Lampp, perhaps) to get

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

All solutions to this homogeneous system are given by $(\frac{1}{2}z, -z, z)$. This is $\mathcal{NS}(A)$. Setting $t = z$, we get the parametric equation of a line $\mathcal{L}$ in $\mathbb{R}^3$:

$x = \frac{1}{2}t, y = -t, z = t$. When $t = 0$, it is obvious that this line passes through the origin, so the line $\mathcal{L}$ is a subspace of $\mathbb{R}^3$ (the result of a previous problem).

One can use Lampp to visualize the intersection of these two subspaces by entering the associated matrix in a 4×4 matrix with zeros in the last two rows and graphing the system. Then one can enter the row reduced form in the bottom two rows and graph the resulting system. Using the Segment button, we can see that the intersection of the four planes is on the line $\mathcal{L}$, as in Figure 5.22.

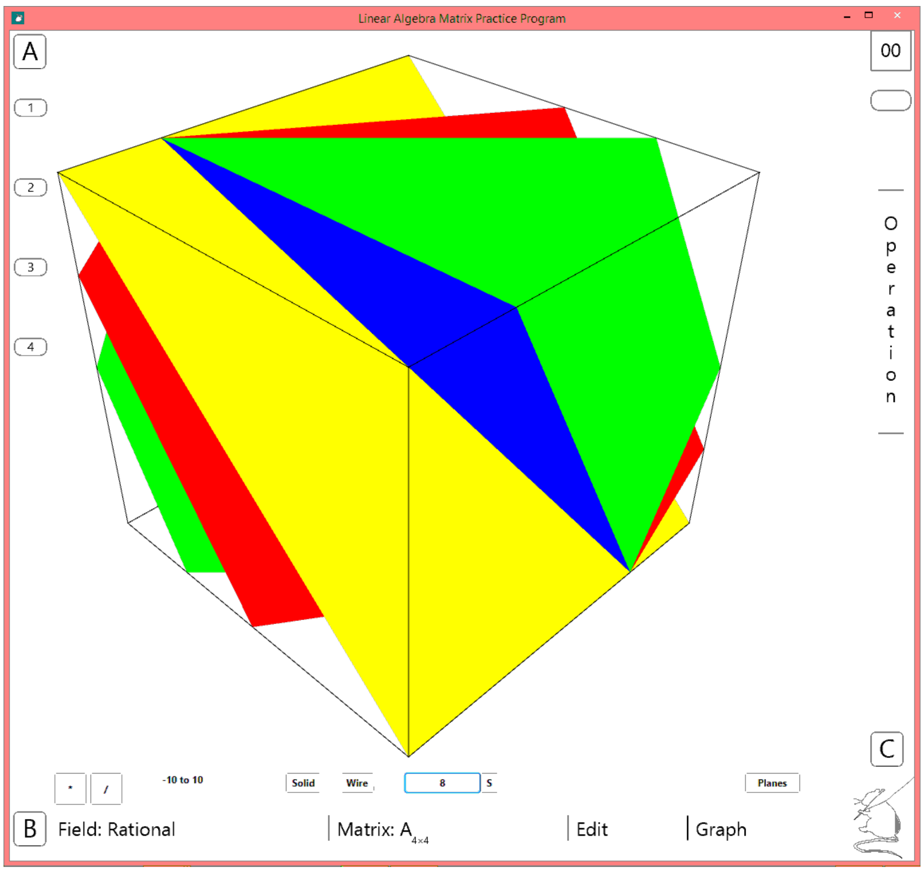


Figure 5.22: Four planes intersecting on one line

Suppose we wanted to find the intersection of two planes which do not go through the origin. We could have, for example, the nonhomogeneous system:

$$\begin{aligned} 2x + 4y + z &= 18 \\ x - y - z &= -16 \end{aligned}$$

To get a particular solution we start by writing the associated matrix.

$$\begin{pmatrix} 2 & 4 & 1 & 18 \\ 1 & -1 & -1 & -16 \end{pmatrix}$$

This matrix can be row reduced to get

$$\begin{pmatrix} 1 & 0 & -\frac{1}{2} & -\frac{23}{3} \\ 0 & 1 & \frac{1}{2} & \frac{25}{3} \end{pmatrix}$$

There are an infinite number of solutions. If we let $z = 0$, then one solution is $(-\frac{23}{3}, \frac{25}{3}, 0)^T$.

The associated homogeneous system reduces to

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

If we let $t = z$, then we have all solutions given by $t(\frac{1}{2}, -\frac{1}{2}, 1)^T$. So, all solutions of the nonhomogeneous system is given as a linear combination

$$X = \begin{pmatrix} -\frac{23}{3} \\ \frac{25}{3} \\ 0 \end{pmatrix} + t \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}$$

which, as we have seen, is a vector equation for a line in $\mathbb{R}^3$. It is interesting to graph all three matrices using Lampp.

Exercises 5.4

1. In $\mathbb{R}^3$, let $\mathcal{S}_1$ be the plane defined by $2x - y - z = 0$, let $\mathcal{S}_2$ be the plane defined by $x + 2y + 3z = 0$ and let $\mathcal{S}_3$ be the plane defined by $5x + z = 0$. Find $\mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_3$.

2. In $\mathbb{R}^3$, let $\mathcal{S}_1$ be the plane defined by $2x - y - z = 0$, let $\mathcal{S}_2$ be the plane defined by $x + 2y + 3z = 0$ and let $\mathcal{S}_3$ be the plane defined by $x + y = 0$. Find $\mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_3$.

3. Find a vector equation for the line of intersection of the following planes (if it exists).

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

5.5 Linear Spans

It would be very convenient to have a method of describing a vector space in such a way that we can find vectors that belong to it. Since a vector space is a very general thing, it is not possible to give a good method that encompasses every vector space. However, there is a method based on the linear combination of vectors that is a very good method of giving a description of all the vectors in certain types of vector spaces. It is called the **span** of a vector space. The word *span* will be used as both a verb and a noun. Each usage will have a distinct, although related, interpretation. We will start with a definition of the verb span.

Definition 5.12. *A finite list of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in a vector space $\mathcal{V}$ spans $\mathcal{V}$ if every vector in $\mathcal{V}$ can be written as a linear combination of these vectors. In other words, if we have $\mathcal{V}$ over $\mathbb{F}$, for every $\vec{v} \in \mathcal{V}$, there are n scalars $a_i \in \mathbb{F}$ such that*

$$\vec{v} = a_1 \vec{v}_1 + a_2 \vec{v}_2 + \ldots + a_n \vec{v}_n$$

We say that a list of vectors spans a vector space. This is slightly different from *the* span of a list of vectors. In this case we use span as a noun.

Definition 5.13. *The span of a finite list of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in a vector space $\mathcal{V}$ over $\mathbb{F}$ is all vectors that can be formed from a finite linear combination of these vectors using scalars in $\mathbb{F}$. Let S stand for a list of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in a vector space $\mathcal{V}$ over $\mathbb{F}$, then* $\mathbf{span}(S)$ *(span of S) are all vectors $\vec{v}$ such that*

$$\vec{v} = a_1 \vec{v}_1 + a_2 \vec{v}_2 + \ldots + a_n \vec{v}_n$$

for any n scalars $a_i \in \mathbb{F}$.

Note that the lists of vectors used in Definition 5.12 and Definition 5.13 are finite, but the number of vectors that are linear combinations of these vectors, $\mathbf{span}(S)$, could be infinite. (It is possible to discuss a vector space with an infinite set of spanning vectors but this text does not.)

Theorem 5.2 from page 119 can easily be used to demonstrate that if S is a collection of vectors from a subspace $\mathcal{V}$, then $\mathbf{span}(S)$ is a subspace of $\mathcal{V}$.

We have already worked with the span of a set (a collection) of vectors. The following example illustrates this.

Suppose we have four vectors $\vec{e}_1, \vec{e}_2, \vec{e}_3$ and $\vec{e}_4$ from $\mathbb{F}^{4\times 1}$ such that

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{e}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

We can say that these vectors span $\mathbb{F}^{4\times 1}$ since any vector in $\mathbb{F}^{4\times 1}$ is of the form

$$\vec{f} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = a_1 \vec{e}_1 + a_2 \vec{e}_2 + a_3 \vec{e}_3 + a_4 \vec{e}_4$$

The vectors $\vec{e}_1, \vec{e}_2, \vec{e}_3$ and $\vec{e}_4$ are often called the standard unit vectors of $\mathbb{F}^{4\times 1}$.

It is possible to find out if a particular vector is in a given span. Suppose we had three vectors in $\mathbb{R}^{1\times 3}$, $\vec{r}_1 = (1, -1, 1)$, $\vec{r}_2 = (3, 7, 8)$ and $\vec{r}_3 = (2, 5, 6)$. Call this collection of vectors S. We would like to know if $\vec{s} = (2, 3, -2)$ is in **span**(S).

To answer this, we need to know if $\vec{s}$ is a linear combination of the vectors that make up S. Do scalars a_1, a_2 and $a_3 \in \mathbb{R}$ exist such that $\vec{s} = a_1\vec{r}_1 + a_1 2\vec{r}_2 + a_3\vec{r}_3$? We can write this as a vector equation.

$$\begin{aligned}\begin{pmatrix} 2 & 3 & -2 \end{pmatrix} &= a_1\begin{pmatrix} 1 & -1 & 1 \end{pmatrix} + a_2\begin{pmatrix} 3 & 7 & 8 \end{pmatrix} + a_3\begin{pmatrix} 2 & 5 & 6 \end{pmatrix} \\ &= \begin{pmatrix} a_1 + 3a_2 + 2a_3 & -a_1 + 7a_2 + 5a_3 & a_1 + 8a_2 + 6a_3 \end{pmatrix}\end{aligned}$$

This last vector is equivalent to the system

$$\begin{aligned} a_1 + 3a_2 + 2a_3 &= 2 \\ -a_1 + 7a_2 + 5a_3 &= 3 \\ a_1 + 8a_2 + 6a_3 &= -2 \end{aligned}$$

To decide if this system has a solution, we row reduce the associated augmented matrix.

$$\begin{pmatrix} 1 & 3 & 2 & 2 \\ -1 & 7 & 5 & 3 \\ 1 & 8 & 6 & -2 \end{pmatrix} \text{ is row equivalent to } \begin{pmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & \frac{4}{5} & -\frac{4}{5} \\ 0 & 0 & -1 & 13 \end{pmatrix}$$

This is enough reduction to show that this system is solvable so $\vec{s} = (2, 3, -2)$ is in **span**(S). If we continue the reduction to row reduced echelon form, we find that $a_1 = -\frac{4}{5}$, $a_2 = \frac{48}{5}$ and $a_3 = -13$ so that

$$\vec{s} = -\frac{4}{5}\vec{r}_1 + \frac{48}{5}\vec{r}_2 - 13\vec{r}_3$$

Sometimes we need to compare two subspaces. We would like to know when they are equal. The following definition gives one way of defining the equality of subspaces.

Definition 5.14. *Two subspaces, $\mathcal{S}_1$ and $\mathcal{S}_2$, in a vector space $\mathcal{V}$ over $\mathbb{F}$, are equal if $\mathcal{S}_1$ is a subspace of $\mathcal{S}_2$ and $\mathcal{S}_2$ is a subspace of $\mathcal{S}_1$.*

A consequence of this definition means that, if $\mathcal{S}_1 = \mathcal{S}_2$, then $\mathcal{S}_1 \cap \mathcal{S}_2 = \mathcal{S}_2 \cap \mathcal{S}_1 = \mathcal{S}_1 = \mathcal{S}_2$. Both $\mathcal{S}_1$ and $\mathcal{S}_2$ contain exactly the same vectors and no others.

Suppose we want to know if the vectors $\vec{r}_1$, $\vec{r}_2$ and $\vec{r}_3$ from the preceding example span $\mathbb{R}^{1\times 3}$. Can the calculations we did be carried out for any $\vec{r}$ in

$\mathbb{R}^{1\times 3}$? In general, a vector $\vec{r}$ in $\mathbb{R}^{1\times 3}$ can be written (r_1, r_2, r_3), where each $r_i \in \mathbb{R}$. This leads to the system

$$\begin{aligned} a_1 + 3a_2 + 2a_3 &= r_1 \\ -a_1 + 7a_2 + 5a_3 &= r_2 \\ a_1 + 8a_2 + 6a_3 &= r_3 \end{aligned}$$

We have already shown that the coefficient matrix for this system is nonsingular and therefore the system has a unique solution for any $\vec{r}$ in $\mathbb{R}^{1\times 3}$. Ergo, $\mathbf{span}(S) = \mathbb{R}^{1\times 3}$.

Since $\mathbb{R}^3$ is Euclidean 3-space, we can have a geometrical interpretation of the span of two vectors in $\mathbb{R}^3$. Let S be the two vectors $\vec{s}_1 = (1, -1, 3)$ and $\vec{s}_2 = (6, 1, 2)$. Then $\mathbf{span}(S)$ is all vectors $\vec{s} = a_1(1, -1, 3) + a_2(6, 1, 2)$ with $a_i \in \mathbb{R}$. In terms of a coordinate system for $\mathbb{R}^3$, we have $x = a_1 + 6a_2$, $y = -a_1 + a_2$ and $z = 3a_1 + 2a_2$. We think of x, y, z as being fixed so we have three equations in the two unknowns a_1, a_2. We solve this using an augmented matrix where the columns refer to coefficients for a_1, a_2, x, y and z, respectively.

$$\left(\begin{array}{rr|rrr} 1 & 6 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 3 & 2 & 0 & 0 & 1 \end{array}\right) \text{ which row reduces to } \left(\begin{array}{rr|rrr} 1 & 0 & \frac{1}{7} & -\frac{6}{7} & 0 \\ 0 & 1 & \frac{1}{7} & \frac{1}{7} & 0 \\ 0 & 0 & 5 & -16 & -7 \end{array}\right)$$

This system is only consistent when $5x - 16y - 7z = 0$, which is an equation of a plane in $\mathbb{R}^3$. If $5x - 16y - 7z \neq 0$, then the system is inconsistent, because $0a_1 + 0a_2$ must equal 0 for any values of a_1 and a_2. There could be no solutions for this system. It is possible to check this answer by substituting the values for $\vec{s}_1$ and $\vec{s}_2$ into the equation for the plane.

Besides the null space, there are two other important subspaces concerning matrices which we will define.

Definition 5.15. *If A is an $m \times n$ matrix over $\mathbb{F}$, then $\mathcal{RS}(A)$, the row space of A, is the subspace spanned by the rows of A.*

Definition 5.16. *If A is an $m \times n$ matrix over $\mathbb{F}$, then $\mathcal{CS}(A)$, the column space of A, is the subspace spanned by the columns of A.*

The column space of a matrix A is related to the idea of **column equivalence** of two matrices. A matrix A is column equivalent to a matrix B if one can make B from A using a finite number of column operations. A column operation on a matrix A can be represented by multiplying one of the three elementary matrices *on the right*. In order to refresh our memory of row equivalence and other theorems from long past pages, we will look a bit at column equivalence.

Definition 5.17. ***Elementary Matrices for Column Operations***

E Type 1 *Interchange any two columns of* $\mathbf{I}$.

E Type ② *Multiply any column of* $\mathbf{I}$ *by a nonzero scalar.*

E Type ③ *Add a scalar multiple of any column of* $\mathbf{I}$ *to a different column.*

If E is an elementary matrix of Type ① and A is a matrix, then AE would have the effect of interchanging two columns of A.

All the theorems involving row operations can be used for column operations by noting that column operations on a matrix A are the same as row operations on A^T. To see this we need to recall two theorems which were assigned as problems. We need that if E is an elementary matrix, then E^T is an elementary matrix. We also need that the transpose of the product of two matrices is the product of the transpose of the second matrix by the first, that $(EA)^T = A^T E^T$. This shows that the transpose of one type of elementary matrix (interchanging two rows, for instance) is the same type of elementary matrix for columns (interchanging two columns). (Can you see why?)

We say that a matrix A is **equivalent** to a matrix D if one can use a finite number of row and/or column operations to create D from A. Since any nonsingular matrix can be made up of the product of elementary matrices, saying that A is equivalent to D means that there must be two nonsingular matrices B and C such that $BAC = D$. The matrix B would be a product of elementary row operation matrices and C would be the product of elementary column operations.

One way to look at this is to use partitioned matrices. Suppose A is an $m \times n$ matrix, $\mathbf{I}_m$ is an $m \times m$ identity matrix and $\mathbf{I}_n$ is an $n \times n$ identity matrix. If A is equivalent to D. we can write two $(m+n) \times (m+n)$ matrices to illustrate this:

$$\left(\begin{array}{c|c} A & \mathbf{I}_m \\ \hline \mathbf{I}_n & \mathbf{0} \end{array}\right) \text{ is equivalent to } \left(\begin{array}{c|c} D & B \\ \hline C & \mathbf{0} \end{array}\right)$$

In other words, $BAC = D$, because multiplying B on the left is equivalent to all the row operations used to change A into D and multiplying C on the right is equivalent to all the column operations used to change A into D.

Please note that if A is row equivalent to B, then A is equivalent to B. Similarly, if A is column equivalent to B, then A is equivalent to B. If A is equivalent to B, we cannot say only from this that A is row equivalent to B. Neither can we conclude that A is column equivalent to B.

We now return to discussions of the column space of a matrix A, $\mathcal{CS}(A)$, by first noting that the columns of an $m \times n$ matrix A over a field $\mathbb{F}$ can be considered as n vectors $\vec{c}_i$ in $\mathbb{F}^{m\times 1}$. A typical vector $\vec{a}$ in $\mathcal{CS}(A)$ would then have the form

$$\vec{a} = x_1\vec{c}_1 + x_2\vec{c}_2 + \cdots + x_n\vec{c}_n$$

where $x_i \in \mathbb{F}$. Using the algebra of matrices, we can rewrite this as

$$\vec{a} = \begin{pmatrix} \vec{c}_1 & \vec{c}_2 & \ldots & \vec{c}_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

This leads us to conclude that

$$\mathcal{CS}(A) = \text{ all matrices } AX \text{ such that } X \in \mathbb{F}^{n\times 1}$$

When doing work with vectors, we could rephrase this, using $\vec{x} \in \mathbb{F}^{n\times 1}$, as

$$\mathcal{CS}(A) = \text{ all vectors } \vec{s} \in \mathbb{F}^{1\times m} \text{ such that } A\vec{x} = \vec{s} \text{ has a solution.}$$

Similarly it can be shown that the row space of a matrix A over $\mathbb{F}$ is

$$\mathcal{RS}(A) = \text{ all matrices } YA \text{ such that } Y \in \mathbb{F}^{1\times m}$$

When we want to prove something about all matrices of a certain type, we start by giving a representative of all these matrices and use this representative to acquire the proof. Since the representative could be any matrix of this type, then the proof must hold for all matrices of this type. This method of proof can be used in many situations and, of course, for more than just matrices.

We can now write two simple, but useful, theorems.

Theorem 5.4. *Given an $m \times n$ matrix A and any matrix B such that BA is defined, then $NS(A)$ is a subspace of $NS(BA)$ and $NS(A) = NS(BA)$ if B^{-1} exists.*

Proof. Suppose X is in $NS(A)$. This means that $AX = \mathbf{0}$. Therefore $(BA)X = \mathbf{0}$ which implies that X is in $NS(BA)$.

If B^{-1} exists, then X in $NS(BA)$ implies that $(BA)X = \mathbf{0}$ which lets us write that $B^{-1}(BA)X = AX = \mathbf{0}$. This shows that X is in $NS(A)$, so that $NS(BA)$ is a subspace of $NS(A)$. Therefore, $NS(A) = NS(BA)$. □

Theorem 5.5. *Given an $m \times n$ matrix A and any matrix B such that BA is defined, then $\mathcal{RS}(BA)$ is a subspace of $\mathcal{RS}(A)$ and $\mathcal{RS}(BA) = \mathcal{RS}(A)$ if B^{-1} exists.*

Proof. We start with the fact that, if B is $r \times m$, $\mathcal{RS}(BA) =$ all matrices $Y(BA)$ such that $Y \in \mathbb{F}^{1\times r}$. The commutative law of matrix multiplication allows us to rewrite this as $\mathcal{RS}(BA) =$ all matrices $(YB)A$ such that $Y \in \mathbb{F}^{1\times r}$. The product YB is a $1 \times m$ matrix and $\mathcal{RS}(A) =$ all matrices ZA such that $Z \in \mathbb{F}^{1\times m}$. Therefore $\mathcal{RS}(BA)$ is a subspace of $\mathcal{RS}(A)$.

If $r = m$ and B^{-1} exists, then from the last result we know that $\mathcal{RS}(A) = \mathcal{RS}(B^{-1}(BA))$ and therefore that $\mathcal{RS}(A)$ is a subspace of $\mathcal{RS}(BA)$. From the definition of equality of subspaces, we have $\mathcal{RS}(BA) = \mathcal{RS}(A)$. □

The span of $\vec{0}$ in $\mathbb{R}^3$ is just the point at the origin. The span of a single vector $\vec{u} \neq \vec{0}$ ($\vec{u}$ *not equal to* the zero vector) in $\mathbb{R}^3$ is all scalar multiples of this vector, $a\vec{u}$. It is simply the line through the origin containing the vector $\vec{u}$. The span of two such vectors, $\vec{u}$ and $\vec{w}$, is any vector in $\mathbb{R}^3$ that is a linear combination of $\vec{u}$ and $\vec{w}$. Such a linear combination determines a plane through the origin as long as $\vec{u}$ is not a scalar multiple of $\vec{w}$. We will examine the situation where we have the span of three or more of these type of vectors in the next section.

Problems 5.5

1. If S is a subspace over $\mathbb{R}$, what is the one case where the number of vectors in $\mathbf{span}(S)$ is finite?

2. Suppose we have vectors of the form $\vec{v} = (a, b)^T$, where $a, b \in \mathbb{Q}_3$. Prove that V, the set of all such vectors $\vec{v}$, is a vector space. Is V finite? Suppose we have a set S containing two vectors from V, $\vec{v}_1 = (1, 2)^T$ and $\vec{v}_2 = (2, 1)^T$. Does $\mathbf{span}(S) = V$?

3. Prove that the transpose of an elementary matrix of one type of row operation corresponds to the same type of column operation.

4. Prove that $\mathcal{RS}(A) =$ all matrices YA such that $Y \in \mathbb{F}^{1\times m}$.

5. Prove that $\mathcal{RS}(A) = \mathcal{CS}(A^T)$.

6. Prove that, given an $m \times n$ matrix A and any matrix C such that AC is defined, then $\mathcal{CS}(AC)$ is a subspace of $\mathcal{CS}(A)$ and $\mathcal{CS}(AC) = \mathcal{CS}(A)$ if C^{-1} exists.

7. Show that the two complex numbers, $1 + i$ and $1 - i$, spans $\mathbb{C}$.

8. Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{v}_{n+1}$ be $(n + 1)$ vectors in a vector space $\mathcal{V}$. Suppose $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ spans $\mathcal{V}$. Prove that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{v}_{n+1}$ also spans $\mathcal{V}$.

9. What is the minimum number of vectors one would need to span $\mathbb{R}^3$? (**HINT**: First find three vectors that span $\mathbb{R}^3$ and show that none of them can be written as a linear combination of the others.)

10. Prove that two polynomials cannot span $\mathbb{P}^2$.

11. Show that $\mathbb{R}^{2\times 2}$ can be spanned by nonsingular matrices.

12. Let S stand for a list of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in a vector space $\mathcal{V}$ over $\mathbb{F}$. If $\vec{u}$ and $\vec{w}$ are in $\mathbf{span}(S)$, prove that $\vec{u} + \vec{w}$ is in $\mathbf{span}(S)$. Prove that $a\vec{u}$, $a \in \mathbb{F}$, is also in $\mathbf{span}(S)$.

13. Let C stand for any vector of the form $\vec{c} = bi$, where $b \in \mathbb{R}$ and $i = \sqrt{-1}$. Show that $\mathbf{span}(C)$ is a proper subspace of $\mathbb{C}$. If R is any vector of the form $\vec{r} = a$, where $a \in \mathbb{R}$, show that $\mathbf{span}(R)$ is a proper subspace of $\mathbb{C}$.

14. Using R and S from the previous problem, show that $\mathbf{span}(R) \cap \mathbf{span}(C) = \vec{0}$. (**HINT**: Assume that there exists a vector $\vec{v} \neq \vec{0}$ in $\mathbf{span}(R) \cap \mathbf{span}(C)$ and use proof by contradiction to show that it cannot exist.)

Exercises 5.5

1. List two vectors that span $\mathbb{R}^2$.

2. List matrices that span $\mathbb{R}^{2\times 2}$.

3. A **monomial** is the product of a scalar and a variable x^n for $n \in \mathbb{Z} > 0$. List the monomials you would need to span $\mathbb{P}^5$.

4. Let $\mathbf{span}(S)$ be the subspace of $\mathbb{R}^3$ spanned by the two vectors $\vec{s}_1 = (1, -1, 3)$ and $\vec{s}_2 = (6, 1, 2)$. Is $(4, 3, -4)$ in this span? What about $(3, 2, -3)$?

5. Let $\mathbf{span}(S)$ be the subspace of $\mathbb{R}^3$ spanned by the three vectors $\vec{s}_1 = (1, -1, 3)$, $\vec{s}_2 = (6, 1, 2)$ and $(4, 3, -4)$. Find an equation for this subspace.

6. Let $\mathbf{span}(S)$ be the subspace of $\mathbb{R}^3$ spanned by the three vectors $\vec{s}_1 = (-1, 1, 3)$, $\vec{s}_2 = (2, 1, 2)$ and $(4, 3, -4)$. Geometrically, what is this subspace?

7. Let $\mathbf{span}(S)$ be the subspace of $\mathbb{C}^2$ spanned by the two vectors $\vec{s}_1 = (-1, i)$, and $\vec{s}_2 = (2 + i, 2)$. Is $(3, 4)$ in this span? What about $(3i, 4i)$? What about $(3, 4i)$?

8. There is a geometrical interpretation of the vector space $\mathbb{C}$. If one has an x, y coordinate system on a plane, one can define the x coordinate to be $\mathbf{span}(R)$, where R is any vector of the form $\vec{r} = a$, for $a \in \mathbb{R}$. Similarly, the y coordinate can be defined to be $\mathbf{span}(C)$, where C stands for any vector of the form $\vec{c} = bi$, where $b \in \mathbb{R}$ and $i = \sqrt{-1}$. Draw this coordinate system. Plot the vectors $\vec{e}_1 = (1, 0)$, $\vec{e}_2 = (0, i)$, $\vec{u} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}i)$ and $\vec{w} = (\frac{4}{5}, \frac{3}{5}i)$. Circle the subspace $\mathbf{span}(R) \cap \mathbf{span}(C)$. Calculate the coordinates for and plot the vectors $\text{proj}_{\vec{e}_1}\vec{u}$, $\text{proj}_{\vec{e}_2}\vec{u}$, $\text{proj}_{\vec{e}_1}\vec{w}$, $\text{proj}_{\vec{e}_2}\vec{w}$, $\text{proj}_{\vec{u}}\vec{w}$ and $\text{proj}_{\vec{w}}\vec{u}$.

5.6 Bases, Linear Independence, Matrix Rank

Suppose we have S, a list of n vectors $\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_n$. Furthermore, suppose $\mathbf{span}(S)$ spans a vector space $\mathcal{V}$. In one of our problems in the preceding section, we have seen that adding more vectors to S still allows us to say that $\mathbf{span}(S)$ spans the vector space $\mathcal{V}$. In order to be conservative of our need to write out vectors, it would be nice to have some way of knowing what is the minimal number of vectors one needs to span a vector space. We can then use this minimum list to write a linear combination of these vectors which we can use as a representative of the vector space $\mathcal{V}$. We are then able to use this representative equation to prove many properties of $\mathcal{V}$.

In order to weed out unnecessary vectors from a list that spans a vector space, we must first have a way of identifying them. We will call these vectors **linearly dependent**.

Definition 5.18. *A collection of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ of a vector space $\mathcal{V}$ over $\mathbb{F}$ are said to be linearly dependent if there exists n scalars x_i in $\mathbb{F}$ such that*

$$\sum_{i=1}^{n} x_i^2 \neq 0 \quad \textit{and} \quad \sum_{i=1}^{n} x_i \vec{v}_i = \vec{0}$$

The first condition about the scalars, that $\sum_{i=1}^{n} x_i^2 \neq 0$, is just another way of saying that they all can't be zero. Another way of saying this, for vectors in the reals, would be that if $\vec{x} = (x_1, x_2, \ldots, x_n)$, then $\vec{x} \cdot \vec{x} > 0$ or $\|\vec{x}\| > 0$.

Any list of vectors $\vec{v}_i$ are called **linearly independent** if they are not linearly dependent. In other words $\vec{v}_i$ are linearly independent when

$$\sum_{i=1}^{n} x_i \vec{v}_i = \vec{0} \text{ is true only if } x_1 = x_2 = \cdots = x_n = 0$$

Suppose we have n vectors $\vec{s}_i$ that span a vector space $\mathcal{S}$ over $\mathbb{F}$. Then any vector $\vec{u}$ in $\mathcal{S}$ may be written as a linear combination of these vectors

$$\vec{u} = \sum_{i=1}^{n} a_i \vec{s}_i \text{ for } a_i \in \mathbb{F}$$

Suppose that we could also write

$$\vec{u} = \sum_{i=1}^{n} c_i \vec{s}_i \text{ for } c_i \in \mathbb{F}$$

where not every $a_i = c_i$. Subtraction gives us

$$\vec{u} - \vec{u} = \sum_{i=1}^{n} (a_i - c_i) \vec{s}_i = \vec{0}$$

From this we can deduce that at least one of the $\vec{s}_i$ may be written as a linear combination of the others. We do this by supposing $a_j - c_j \neq 0$. Then

$$-(a_j - c_j)\vec{s}_j = \sum_{i=1}^{j-1} (a_i - c_i)\vec{s}_i + \sum_{i=j+1}^{n} (a_i - c_i)\vec{s}_i$$

If the linear combination

$$\vec{u} = \sum_{i=1}^{n} a_i \vec{s}_i \text{ for } a_i \in \mathbb{F}$$

is composed of linearly independent vectors $\vec{s}_i$, then each a_i is *unique*. We cannot have

$$\vec{u} = \sum_{i=1}^{n} c_i \vec{s}_i \text{ for } c_i \in \mathbb{F}$$

where not every $a_i = c_i$.

A linearly independent list of vectors from a vector space $\mathcal{V}$ which spans $\mathcal{V}$ is an important idea. It is so important that we give it it's own name. We call it a **basis**.

Definition 5.19. *A basis for a vector space $\mathcal{V}$ is a collection of linearly independent vectors from $\mathcal{V}$ which spans $\mathcal{V}$.*

From our previous discussion, we have seen that the scalar multiples of the vectors in a basis are unique, since the scalar multiples of any independent linear combination are unique. Of course, different vectors may form a basis for the same vector space. Also, as we noted earlier, not all vector spaces contain a basis as we have defined it.

One frequently occurring problem is to find out if a collection of vectors from a vector space forms a basis. We need to have some method of determining if the vectors are linearly independent.

As luck would have it, we can define an algorithm that will determine whether a list of vectors is a basis or not. It involves the solution of a homogeneous system of linear equations. We have already done this many times. We simply represent the system of linear equations in terms of a matrix of coefficients A and a matrix of variables X and then solve $AX = \mathbf{0}$ using Gauss-Jordan elimination.

Before we give this algorithm, let us return to the central questions we posed when we first introduced systems of equations. How is linear algebra used to discover if solutions to systems of equations exist? If a solution exists, is it unique? How do we find all the solutions? We are just about in a position to answer all of these questions, which directly pertains to using Gauss-Jordan elimination to find out if certain vectors from a vector space form a basis.

We have seen that we can encounter systems of equations for which $AX = B$ has no solutions. We have said that these types of nonhomogeneous systems are inconsistent. Remembering our previous definition of rank, we can directly say that $AX = B$ is consistent if the rank of A is equal to the rank of the augmented matrix $(A|B)$. Our previous definition of rank was useful but not specific enough. We will strengthen the definition using some nice properties of bases of vector spaces and linearly independent vectors and by introducing the concept of the **dimension** of a vector space. We will start with the **Exchange Theorem**.

Theorem 5.6. *Suppose that a collection of n vectors $\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_n$ of a vector space $\mathcal{V}$ over $\mathbb{F}$ spans $\mathcal{V}$. Then any collection of linearly independent vectors from $\mathcal{V}$ must have $m \leq n$ vectors.*

Proof. Suppose that $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m$ is a collection of independent vectors from $\mathcal{V}$ where $m > n$. Then we can break this into two collections of vectors. One collection of n vectors, any of which we will represent by $\vec{w}_j$ (where $1 \leq j \leq n$) and another collection of $m - n$ vectors for which we will use $\vec{w}_k$ (where $n + 1 \leq k \leq m$) to represent any one of them.

We know that any $\vec{w}_j$ is a linear combination of the vectors $\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_n$. Therefore, we can immediately state that the collection $\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_n, \vec{w}_1$ is linearly dependent and that it also spans $\mathcal{V}$ (this is a result from a problem from the previous section). From the definition of span and linear dependence, we know there are unique a_i and a unique c_1, not all zero, such that

$$\vec{0} = c_1\vec{w}_1 + \sum_{i=1}^{n} a_i\vec{s}_i$$

All the a_i cannot be zero, otherwise we would have $c_1 = 0$ and our collection would be linearly independent. By rearranging terms, if necessary, we can assume that $a_1 \neq 0$. Therefore $\vec{s}_1$ is dependent on the rest. Therefore we can remove it and still have that the collection $\vec{w}_1, \vec{s}_2, \ldots, \vec{s}_n$ spans $\mathcal{V}$. Since $1 \leq j \leq n$ and $1 \leq i \leq n$, we can continue in this way until we have exchanged all $\vec{s}_i$ for $\vec{w}_i$ and have that the collection $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n$ spans $\mathcal{V}$. (Notice that every time we introduce a new $\vec{w}_j$ into the collection, it must only be a linear combination of the remaining $\vec{s}_i$, since the $\vec{w}_j$ is independent of the $(n-j-1)$ vectors, $\vec{w}_i$, we already introduced.) Then any $\vec{w}_k$ (where $n+1 \leq k \leq m$) must be a linear combination of these vectors so, therefore, dependent. This contradicts our assumption of $m > n$ independent vectors. We can then say that m cannot be greater than n. □

This remarkable result leads directly to a proof of the following theorem.

Theorem 5.7. *If a vector space has a basis composed of a finite number of vectors, then any other basis for that vector space contains exactly the same number of vectors.*

Proof. Let A stand for a collection of n vectors that form a basis for $\mathcal{V}$. Let B stand for m vectors that form a basis for $\mathcal{V}$. Then

1. $\mathbf{span}(A) = \mathcal{V}$

2. The vectors in B are linearly independent.

By the previous exchange theorem, $n \leq m$. Similarly,

1. $\mathbf{span}(B) = \mathcal{V}$

2. The vectors in A are linearly independent.

Again, by the exchange theorem, $m \leq n$. Therefore, $m = n$. □

This theorem allows us to make the following definition which allows us to classify certain vector spaces.

Definition 5.20. *If $\mathcal{V}$ is a finite dimensional vector space, then dim$(\mathcal{V})$, the dimension of $\mathcal{V}$, is the number of vectors in a basis for $\mathcal{V}$.*

So if a basis for a vector space $\mathcal{S}$ contained 4 vectors, then we would call $\mathcal{S}$ a 4-dimensional vector space or say that $\mathcal{S}$ has four dimensions. The classic example of a vector space that does not have a basis, as we define it, is $\mathbb{P}$, the vector space of polynomials of any degree. It is called infinite dimensional.

We will now show that for any matrix, the dimension of its row space is equal to the dimension of its column space. We will redefine the rank of a matrix to be this common dimension and then show how the rank may be used to characterize all solutions of a system $AX = B$. However, we have a bit of work ahead of us before we get there. We begin with a theorem about the dimension of a subspace of a finite dimensional vector space.

Theorem 5.8. *If $\mathcal{S}$ is a subspace of a finite dimensional vector space $\mathcal{V}$, then*

$$dim(\mathcal{S}) \leq dim(\mathcal{V})$$

Proof. Let $dim(\mathcal{V}) = n$. Any collection of linearly independent vectors in $\mathcal{S}$ must also be a collection of linearly independent vectors in $\mathcal{V}$. By the exchange theorem and the fact that $\mathcal{V}$ contains a basis composed of n linearly independent vectors that span $\mathcal{V}$, any set of linearly independent vectors in $\mathcal{S}$ can contain at most n vectors. Therefore $\mathcal{S}$ is finite dimensional and $dim(\mathcal{S}) \leq n$ because any basis in $\mathcal{S}$ is a set of linearly independent vectors. □

The preceding theorem lets us prove the following.

Theorem 5.9. *Any n linearly independent vectors in a vector space $\mathcal{S}$ of dimension n are a basis for $\mathcal{S}$.*

Proof. Let V be a set $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ of n linearly independent vectors in $\mathcal{S}$. If they span $\mathcal{S}$, then they form a basis. If not, then there is a vector $\vec{w}$ in $\mathcal{S}$ that is not in $\mathbf{span}(V)$. Let W be the set of $n+1$ linearly independent vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{w}$. We can say they are linearly independent because, if

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n + a_{n+1}\vec{w} = 0$$

then $a_{n+1} = 0$, otherwise we could write $\vec{w}$ as a linear combination of the other vectors (How?). Therefore, if $a_{n+1} = 0$, then

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n = 0$$

which means all $a_i = 0$ since the $\vec{v}_i$'s are linearly independent. Since all the vectors in W are in $\mathcal{S}$, then $\mathbf{span}(W)$ is a subspace of $\mathcal{S}$. Also, since all the vectors in W are linearly independent, they form a basis for $\mathbf{span}(W)$. This means $dim(\mathcal{SP}(W)) = n+1$. This contradicts the previous theorem, which says that $dim(\mathcal{SP}(W)) \leq n$. There can be no vector $\vec{w}$ in $\mathcal{S}$ that is not in $\mathbf{span}(V)$. Thus, the vectors in V span $\mathcal{S}$ and therefore satisfy the criteria for being a basis for $\mathcal{S}$. □

The reader should assure themselves that this theorem allows us to say that we can always extend a set of linearly independent vectors in a finite dimensional vector space into a basis by adding zero or more linearly independent vectors. Some thinking also allows one to see that almost any set of linearly dependent vectors that span a vector space contains a smaller set of linearly independent vectors that span the vector space. (Can you think of the one exception?)

We have two more major theorems to prove before we can start arranging our knowledge in order to define an algorithm to compute all solutions of a system $AX = B$ and the necessary and sufficient conditions for solutions to exist. The following theorem establishes an important relationship between the number of columns in a matrix and the dimension of its column space and the dimension of its null space.

Theorem 5.10. *For any $m \times n$ matrix A over $\mathbb{F}$, the number of columns of A equals the sum of the dimensions of the column space of A and the null space of A. That is*

$$n = dim(\mathcal{NS}(A)) + dim(\mathcal{CS}(A))$$

Proof. Let $s = dim(\mathcal{NS}((A))$. The null space of A is all vectors X in $\mathbb{F}^{n\times 1}$ such that $AX = \mathbf{0}$. This means that $\mathcal{NS}((A)$ is a subspace of the vector space $\mathbb{F}^{n\times 1}$. Thus we have that $s \leq n$. If $n = s + t$, we need to prove that $t = dim(\mathcal{CS}(A))$. Since $\mathbb{F}^{n\times 1}$ is a finite dimensional vector space, we can have a basis for $\mathcal{NS}(A)$ with at most s vectors. Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_s$ be this basis. We can extend this basis by adding zero or more linearly independent vectors from $\mathbb{F}^{n\times 1}$ to get a basis for $\mathbb{F}^{n\times 1}$. Let $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_t$ be these vectors. We now have that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_s, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_t$ is a basis for $\mathbb{F}^{n\times 1}$. When X is in $\mathbb{F}^{n\times 1}$, then any matrix product AX is in the column space of A. So, for any $\vec{c}$ in $\mathcal{CS}(A)$, we can find an X such that $\vec{c} = AX$. Using the basis we defined for $\mathbb{F}^{n\times 1}$, then

$$X = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_s\vec{v}_s + c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_t\vec{w}_t$$

for some scalars a_i and c_i in $\mathbb{F}$. Using the distributive law for matrix multiplication and summation notation, we can write $\vec{c} = AX$ as

$$\vec{c} = AX = \sum_{i=1}^{s} a_i A\vec{v}_i + \sum_{i=1}^{t} c_i A\vec{w}_i$$

Since the $\vec{v}_i$ form a basis for $\mathcal{NS}(A)$, then

$$\sum_{i=1}^{s} a_i A\vec{v}_i = \mathbf{0}$$

and then

$$\vec{c} = AX = \sum_{i=1}^{t} c_i A\vec{w}_i$$

We have said that $\vec{c}$ is any vector in $\mathcal{CS}(A)$, so the t vectors of the form $A\vec{w}_i$ span $\mathcal{CS}(A)$. If we can show that these vectors are linearly independent, then they form a basis for $\mathcal{CS}(A)$. This will show that $dim(\mathcal{CS}(A)) = t$ and complete the proof.

Let us suppose that we have $b_i \in \mathbb{F}$ such that

$$\sum_{i=1}^{t} b_i A\vec{w}_i = \mathbf{0}$$

which can be rewritten as

$$A\sum_{i=1}^{t} b_i \vec{w}_i = \mathbf{0}$$

This means that

$$\vec{u} = \sum_{i=1}^{t} b_i \vec{w}_i$$

is in $\mathcal{NS}(A)$. For some $d_i \in \mathbb{F}$, we can write $\vec{u}$ in terms of our basis for $\mathcal{NS}(A)$.

$$\vec{u} = \sum_{i=1}^{s} d_i \vec{v}_i$$

Subtracting one form from the other gives

$$\vec{u} - \vec{u} = \sum_{i=1}^{s} d_i \vec{v}_i - \sum_{i=1}^{t} b_i \vec{w}_i = \mathbf{0}$$

But we have seen that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_s, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_t$ is a basis for $\mathbb{F}^{n\times 1}$. This means $b_1 = b_2 = \cdots = b_t = d_1 = d_2 = \cdots = d_s = 0$, from the definition of linear independence. We can conclude that the vectors $A\vec{w}_i$ are linearly independent and we have completed our proof. $\square$

We have one more very important theorem with a long, but easy to follow, proof.

Theorem 5.11. *For any $m \times n$ matrix A over $\mathbb{F}$, the dimension of the column space of A equals the dimension of the row space of A. That is*

$$dim(\mathcal{CS}(A)) = dim(\mathcal{RS}(A))$$

Proof. Let A be a $m \times n$ matrix A over $\mathbb{F}$ with $dim(\mathcal{RS}(A)) = r$. If the set of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r$ is a basis for $\mathcal{RS}(A)$, then each row i of A can be written as a row vector, $\vec{r}_i$ in $\mathbb{F}^{1\times n}$, with $1 \leq i \leq m$, and as a linear combination of the basis vectors, for some $c_{i,j} \in \mathbb{F}$.

$$\vec{r}_i = \sum_{j=1}^{r} c_{i,j} \vec{v}_j = (a_{i,1}, a_{i,2}, \ldots, a_{i,n})$$

As well, each $\vec{v}_j$ is in $\mathbb{F}^{1\times n}$ so we can write, for $b_{j,i} \in \mathbb{F}$,

$$\vec{v}_j = (b_{j,1}, b_{j,2}, \ldots, b_{j,n})$$

So $c_{i,j}\vec{v}_j = (c_{i,j}b_{j,1}, c_{i,j}b_{j,2}, \ldots, c_{i,j}b_{j,n})$, which gives the vector equation for each of the m rows of A as

$$\vec{r}_i = \left(\sum_{j=1}^{r} c_{i,j}b_{j,1}, \sum_{j=1}^{r} c_{i,j}b_{j,2}, \ldots, \sum_{j=1}^{r} c_{i,j}b_{j,n}\right) = (a_{i,1}, a_{i,2}, \ldots, a_{i,n})$$

Equating entries gives us, with $1 \leq k \leq n$ and $1 \leq i \leq m$,

$$a_{i,k} = \sum_{j=1}^{r} c_{i,j}b_{j,k}$$

We use this to write a system of m equations

$$\begin{aligned}
a_{1,k} &= c_{1,1}b_{1,k} + c_{1,2}b_{2,k} + \cdots + c_{1,k}b_{r,k} \\
a_{2,k} &= c_{2,1}b_{1,k} + c_{2,2}b_{2,k} + \cdots + c_{2,k}b_{r,k} \\
&\vdots \\
a_{m,k} &= c_{m,1}b_{1,k} + c_{m,2}b_{2,k} + \cdots + c_{m,k}b_{r,k}
\end{aligned}$$

We can then write column k of the matrix A as a column vector and as a linear combination of r vectors

$$\vec{c}_k = \begin{pmatrix} a_{1,k} \\ a_{2,k} \\ \vdots \\ a_{m,k} \end{pmatrix} = b_{1,k}\begin{pmatrix} c_{1,1} \\ c_{2,1} \\ \vdots \\ c_{m,1} \end{pmatrix} + b_{2,k}\begin{pmatrix} c_{1,2} \\ c_{2,2} \\ \vdots \\ c_{m,2} \end{pmatrix} + \cdots + b_{r,k}\begin{pmatrix} c_{1,r} \\ c_{2,r} \\ \vdots \\ c_{m,r} \end{pmatrix}$$

Since $\vec{c}_j$ is any column vector of A, then we have r vectors that span the column space. This implies that the $dim(\mathcal{CS}(A)) \leq r$ and since $r = dim(\mathcal{RS}(A))$, then

$$dim(\mathcal{CS}(A)) \leq dim(\mathcal{RS}(A))$$

We may repeat the same argument using the transpose of A to obtain

$$dim(\mathcal{CS}(A^T)) \leq dim(\mathcal{RS}(A^T))$$

The row space of A^T is the column space of A and the row space of A is the column space of A^T. So we must have that

$$dim(\mathcal{RS}(A)) \leq dim(\mathcal{CS}(A))$$

Since $dim(\mathcal{CS}(A)) \leq dim(\mathcal{RS}(A))$ and $dim(\mathcal{RS}(A)) \leq dim(\mathcal{CS}(A))$, then we complete the proof.

$$dim(\mathcal{CS}(A)) = dim(\mathcal{RS}(A))$$

$\square$

We now give a proper definition of the rank of a matrix.

Definition 5.21. *For any $m \times n$ matrix A over $\mathbb{F}$, the rank of A is given by $rank(A) = dim(\mathcal{RS}(A)) = dim(\mathcal{CS}(A))$*

The dimension of the null space of an $m \times n$ matrix A is often written as $nullity(A)$. We can now rewrite the result of Theorem 5.10 as

$$n = nullity(A) + rank(A)$$

We now have a wealth of theorems which we can use to explore.

Recall that $\mathbf{I}_n$ is an $n \times n$ identity matrix. It should now be obvious that the rows of $\mathbf{I}_n$ are linearly independent and span the subspace $\mathbb{F}^{1\times n}$. Similarly, the columns of $\mathbf{I}_n$ span $\mathbb{F}^{n\times 1}$. This means that $rank(\mathbf{I}_n) = n$. Obviously $nullity(\mathbf{I}_n) = 0$, which means that the null space of $\mathbf{I}_n$ is a zero vector.

We proved that if B is a nonsingular matrix and BA exists, then $\mathcal{CS}(BA) = \mathcal{CS}(A)$. Immediately we have that $rank(BA) = rank(A)$. Similarly, if C is nonsingular and AC exists, we can say that $rank(AC) = rank(A)$. So if C^{-1} and B^{-1} exist, then $rank(BAC) = rank(A)$.

We can also see that any $m \times n$ matrix A is equivalent to a matrix R of the form

$$R = \left(\begin{array}{c|c} \mathbf{I}_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right)$$

where $r = rank(A)$. (Why is this?) Using the results of this discussion, we can show that

Theorem 5.12. *Given two $m \times n$ matrices A and B, then A is equivalent to B if and only if $rank(A) = rank(B)$.*

The next theorem sums up our findings about square matrices. The proof can be obtained by carefully following the results of established theorems.

Theorem 5.13. *Given an $n \times n$ matrix A over $\mathbb{F}$, any of following statements implies the others.*

1. $rank(A) = n$
2. *The row vectors of A are linearly independent.*
3. $\mathcal{RS}(A) = \mathbb{F}^{1\times n}$
4. *The column vectors of A are linearly independent.*
5. $\mathcal{CS}(A) = \mathbb{F}^{n\times 1}$
6. *A^{-1} exists.*

This next theorem sums up the information we will need to find if a linear system $AX = B$ has a solution. This is the theorem we will need to (finally) develop an algorithm for finding solutions to $AX = B$, if they exist.

Theorem 5.14. *Suppose we have $A_{m\times n}$ and a linear system $AX = B$. All of the following statements are true or they are all false.*

1. *$AX = B$ has a solution.*
2. *B is in $\mathcal{CS}(A)$*
3. *$rank(A) = rank(A|B)$, where $A|B$ is the augmented matrix made from A and B.*

Proof. Recalling that $\mathcal{CS}(A)$ is any matrix B in $\mathbb{F}^{m\times 1}$ such that $AX = B$ is solvable, the first two statements are obviously equivalent. Since the number of independent columns of $A|B$ is equal to the rank of $A|B$, we can see that the second statement implies the third. To see that the third statement implies the second, notice that the only way we can add another column B to A to get $(A|B)$ and have that $rank(A) = rank(A|B)$ is if B is in $\mathcal{CS}(A)$. We have shown that any of the three statements implies the others, so that they are all true or all false. $\square$

Again, do not confuse the augmented matrix $A|B$ with the product of two matrices AB.

Let us now look at a scheme to find solutions of a linear system $AX = B$ where A is in $\mathbb{F}^{m\times n}$ and B is in $\mathbb{F}^{m\times 1}$. (This means, of course, that X is in $\mathbb{F}^{n\times 1}$.) We divide up our scheme into different tasks.

Task 1. Determining if a given B is a solution to $AX = B$.

Evaluate $rank(A)$ and $rank(A|B)$. If they are not equal, then there are no solutions for this system (see Theorem 5.14). The easiest way to find out if they are equal is to create an $m \times (n + 1)$ augmented matrix, $A|B$, and use row operations to reduce it to reduced row echelon form. If there is a row of all zeroes except for the column $(n + 1)$ entry, then the ranks are *not* equal. The system is inconsistent and B is not a solution to the system. If the ranks are equal, then B is in $\mathcal{CS}(A)$ and is a solution to the system. Furthermore, if $m = n = rank(A)$, then A is invertible and B is the only solution (recall Theorem 5.13.)

Task 2. Finding all B such that B is a solution to $AX = B$.

The following procedure is an extended form of what we did in Task 1. First, instead of reducing $A|B$, we will reduce $A|C$ where C is an $m \times m$ matrix such that C is zero except for its diagonal entries, which are $c_{i,i} = b_i$. This is easy to do with Lampp. Then use row operations to reduce $A|C$ to reduced row echelon form. If there any rows with only zero entries except in the column entries with indices $> n$, then the system is inconsistent. This is the same information we had before, but now we have the bonus of being able to say why it is inconsistent and for what values of b_i it can be made consistent.

For example, let us examine the system

$$\begin{aligned} -x_1 - 2x_2 - x_3 - \ 5x_4 &= b_1 \\ 2x_1 + 4x_2 + x_3 + \ 8x_4 &= b_2 \\ 4x_1 + 8x_2 \ + 12x_4 &= b_3 \end{aligned}$$

This gives us a matrix equation $AX = B$ which we can also represent as $AX = \mathbf{I}B$. We do this for reasons which will become apparent as we proceed. We write $A|\mathbf{I}_m$ for this system as the augmented matrix

$$\left(\begin{array}{rrrr|rrr} -1 & -2 & -1 & -5 & 1 & 0 & 0 \\ 2 & 4 & 1 & 8 & 0 & 1 & 0 \\ 4 & 8 & 0 & 12 & 0 & 0 & 1 \end{array}\right)$$

Each column of $\mathbf{I}_m$ represents the coefficients for b_i. We row reduce this matrix to reduced echelon form to get

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

The only way that this system can be consistent is if the bottom row is all zero so that $rank(A|\mathbf{I}_m) = 2 = rank(A)$. This happens when $-4b_1 - 4b_2 + b_3 = 0$. So any B which has entries that satisfy this equation will have a solution for $AX = B$. It is possible to find an X which makes $AX = B$ true. We can have a solution space.

For example, if $b_1 = 1$,$b_2 = -2$ and $b_3 = -4$, then the system

$$\begin{aligned} -x_1 - 2x_2 - x_3 - \ 5x_4 &= 1 \\ 2x_1 + 4x_2 + x_3 + \ 8x_4 &= -2 \\ 4x_1 + 8x_2 \ + 12x_4 &= -4 \end{aligned}$$

has a solution. We can immediately go to the reduced row echelon representation

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

using the equations obtained from our previous row reduction, that is

$$\begin{aligned} 1(1) + 1(-2) &= -1 \\ -2(1) - 1(-2) &= 0 \\ -4(1) - 4(-2) + 1(-4) &= 0 \end{aligned}$$

Using parameters, we let $x_2 = s$ and $x_4 = t$ so that any solution is of the form

$$X = \begin{pmatrix} -2s - 3t - 1 \\ s \\ -2t \\ t \end{pmatrix}$$

Task 3. Finding all X such that $AX = \mathbf{0}$.

If $X = \mathbf{0}$, then this is a solution. It is called the **trivial solution**. We are interested in finding X when $X \neq \mathbf{0}$. We have already seen one method using parameters. We will show an alternate method. We have to find a basis for $\mathcal{NS}(A)$. We begin by adding an identity matrix to the bottom of A to get the augmented matrix

$$\left(\frac{A}{\mathbf{I}_n}\right)$$

and then use *column* operations to reduce the augmented matrix to **reduced column echelon form** . A matrix A is in reduced column echelon form if

(a) Any columns consisting entirely of 0's appear at the right of the matrix.

(b) For any column, the leading entry (if it exists) is a 1 and any row containing a leading entry has zeros for entries everywhere else in the row.

(c) For any two consecutive columns with leading entries, the leading entry of the column to the right is below the leading entry of the column to its left.

So for our example we write

$$\left(\begin{array}{rrrr} -1 & -2 & -1 & -5 \\ 2 & 4 & 1 & 8 \\ 4 & 8 & 0 & 12 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \begin{array}{l}\text{column} \\ \text{equivalent to}\end{array} \left(\begin{array}{rrrr} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{8}{3} & -\frac{5}{3} & \frac{2}{3} & \frac{4}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \end{array}\right)$$

The two columns of the bottom portion of the reduced, augmented matrix which do not correspond to leading entries in the top portion are two independent vectors which span $\mathcal{NS}(A)$. Therefore, any X which satisfies $AX = \mathbf{0}$ must be a linear combination of these two vectors. For arbitrary $s, t \in \mathbb{F}$

$$X = s\begin{pmatrix} 1 \\ 0 \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix} + t\begin{pmatrix} 0 \\ 1 \\ \frac{4}{3} \\ -\frac{2}{3} \end{pmatrix}$$

As we said, $s, t \in \mathbb{F}$ are arbitrary, so we could have also writen

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

which looks nicer.

Note: We could just have easily started with the reduced row echelon form of A which we calculated in Task 2 as the top matrix in our augmented matrix. (Why?)

Task 4. Finding all X such that $AX = B$.

We begin by adding an identity matrix and a column of zeros to the bottom of $A|B$ to get the augmented matrix

$$\left(\begin{array}{c|c} A & B \\ \hline \mathbf{I}_n & \mathbf{0} \end{array}\right)$$

and then follow the procedure outlined in Task 1 to determine if a solution is possible. If $m = n = rank(A)$, then there is only one solution, which we can read off directly from the reduced row echelon equivalent matrix we obtained from $A|B$. If an infinite number of solutions are possible, where $rank(A) < n$, we can then continue as we did in Task 3 and find a basis for the null space of A.

With this basis and one particular solution, we can then define all solutions.

For our example we have $B = (1, -2, -4)^T$ and

$$\left(\begin{array}{rrrr|r} -1 & -2 & -1 & -5 & 1 \\ 2 & 4 & 1 & 8 & -2 \\ 4 & 8 & 0 & 12 & -4 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

which we can reduce the top 3 rows and first 4 columns to

$$\left(\begin{array}{rrrr|r} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ -\frac{2}{3} & 1 & 2 & 4 & 0 \\ \frac{1}{3} & 0 & -1 & -2 & 0 \end{array}\right)$$

Letting $x_3 = s$ and $x_4 = t$, if we have $s = t = 0$, then $x_1 = -1$ and $x_2 = 0$, which gives us one solution $X = (-1, 0, 0, 0)^T$. Therefore, any solution X is of the form

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

for any $s, t \in \mathbb{F}$.

Notice that this is the same solution as the one we found using parameters in Task 2.

There are two echelon forms for a matrix, one for rows and one for columns. Furthermore we make the distinction between **echelon form** and **reduced echelon form**. When talking about row echelon form, we may have leading entries in a row that are not equal to one and other entries in its column which may not be zero. Respective examples are given below.

$$\text{row echelon form} \begin{pmatrix} 1 & 3 & 3 & 3 & 8 \\ 0 & 0 & 2 & 2 & 4 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \text{then reduced} \begin{pmatrix} 1 & 3 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Columns 3 and 5 in the reduced row echelon matrix on the right do not have leading ones. These kind of columns are called **free** columns because, if we were to associate a variable from a system of equations with this column, we could represent that variable with a parameter to which we can freely give any value. In other words, our parameters could be $x_3 = s$ and $x_5 = t$.

While we are discussing the reduced row echelon form, this would be a good time to prove one of its features.

Theorem 5.15. *The reduced row echelon form of a matrix is unique.*

Proof. For this proof, we will use the **prime** symbol ($'$) to denote a certain kind of matrix, as well as three indices, $m, r, s \in \mathbb{Z}^+$, with $r + 1 = s$. If $M_{m\times s}$ is a matrix, we define $M'_{m\times r}$ (read M prime m by r) to be the $m \times r$ matrix created by removing the s^{th} column from the right end of M.

We will use induction to prove our theorem. Suppose we have a matrix A with only one column. If A is the zero matrix, $\mathbf{0}$, then A is in row reduced echelon form and only equivalent to itself. If A has a non-zero entry and is in row reduced echelon form, then A must be equal to the $m \times 1$ column matrix $(1, 0, \ldots, 0)^T$. This makes up our first induction step.

For our second induction step, we assume that for every size matrix up to and including $A'_{m\times r}$ there is a unique $R'_{m\times r}$ that is in row reduced echelon form. Make $A_{m\times s}$ by adding a column to the end of A'. If we show that A has a unique row reduced echelon form, $R_{m\times s}$, then we will have completed our proof by induction.

We first make an important observation. If $A_{m\times s}$ can be row reduced to two row reduced echelon matrices, $B_{m\times s}$ and $C_{m\times s}$, then B' and C' are also in row reduced echelon form and row equivalent to A'. We also can say, from our step two induction assumption, that $B' = C'$.

This means either $B = C$ or, if $B \neq C$, that B and C differ in their last columns. We desire to prove that the last columns must be equal, so we will assume $B \neq C$ is true and use proof by contradiction.

Since $B \neq C$, B and C must only differ in the s^{th} column. This assures us that the i^{th} row of B (for some specific integer i) is not equal to the i^{th} row of C. Let $\vec{u} = (u_1, \ldots, u_s)^T$ be a column vector such that $B\vec{u} = \vec{0}$. Since B and C are row equivalent, we also have $C\vec{u} = \vec{0}$, as $\vec{u}$ is a solution to the two associated homogeneous systems for $B\vec{u} = \vec{0}$ and $C\vec{u} = \vec{0}$. So we also have $(B - C)\vec{u} = \vec{0}$.

The first r columns of $(B - C)$ are zero columns, since $B' = C'$. This means there must be some index $i < m$ such that $b_{i,s} \neq c_{i,s}$. Thus the i^{th} entry of $(B - C)\vec{u}$ is $(b_{i,s} - c_{i,s})u_s$ and so we must have $u_s = 0$ and the last columns of B and C contain leading ones since otherwise they would be free columns and the value of u_s could be arbitrarily selected. However, the first r columns of B and C are identical, so the row with the leading entry equal to one must be the same for both B and C. This is the first zero row of the reduced echelon form of A'. A contradiction arises since all other entries in the s^{th} columns of B and C have to be zero.

$\square$

The preceding theorem shows that we can determine whether two matrices are row equivalent by evaluating their row reduced echelon forms. The matrices are row equivalent if and only if they have the same row reduced echelon form.

Problems 5.6

1. Prove that any set (any list) of vectors containing the zero vector is linearly dependent.

2. Show that a set containing just one non-zero vector is linearly independent.

3. Prove that if S refers to a set of vectors such that $\textbf{span}(S) = \mathcal{V}$, then S must contain a basis for $\mathcal{V}$, provided that $\mathcal{V}$ does not simply contain only a zero vector.

4. Prove that, if A is $m \times n$, then $m = \mathit{nullity}(A^T) + \mathit{rank}(A)$.

5. Show that $\mathit{rank}(AB) \leq \mathit{rank}(A)$ and also that $\mathit{rank}(AB) \leq \mathit{rank}(B)$.

6. Prove that for any two $m \times n$ matrices A and B such that A is equivalent to B, then $\mathit{rank}(A) = \mathit{rank}(B)$.

7. Prove that for any two $m \times n$ matrices A and B such that $\mathit{rank}(A) = \mathit{rank}(B)$, then A is equivalent to B. Combined with the result of the previous problem, this proves Theorem 5.12.

8. Show that $\mathit{rank}(A) = \mathit{rank}(A^T)$ for any matrix A over a field $\mathbb{F}$.

9. What would be a basis for the vector space composed of all diagonal $n \times n$ matrices? What is the dimension of this vector space?

10. Suppose we are given a vector space $\mathcal{V}$ of dimension n and we also have m independent vectors from this vector space, $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$, such that $m < n$. Show that we can add $n - m$ vectors to this set of vectors to create a basis for $\mathcal{V}$.

11. If we have n vectors which span a vector space $\mathcal{V}$, show that $dim(\mathcal{V}) \leq n$.

12. Suppose that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ form a basis for a vector space $\mathcal{V}$. Let

$$\vec{u}_i = \sum_{j=1}^{i} \vec{v}_j$$

for $1 \leq i \leq n$. Show that the set containing the vectors $\vec{u}_i$ is also a basis for $\mathcal{V}$.

13. Show that the reduced column echelon form of a matrix is unique. (This should be easy!)

14. Demonstrate that the row echelon form of a matrix is not unique.

15. Let $A \in \mathbb{F}^{m\times n}$. Prove there is a unique $R \in \mathbb{F}^{m\times n}$ where R is in row reduced echelon form and $\mathcal{RS}(A) = \mathcal{RS}(R)$.

Exercises 5.6

1. Show that the given lists of vectors over $\mathbb{R}$ are dependent or independent.

(a) $(1,2)^T$
$(0,1)^T$
$(2,3)^T$

(b) $(1,3,2)^T$
$(3,-1,2)^T$
$(3,1,2)^T$

(c) $(1,3,3)^T$
$(7,-1,5)^T$
$(3,-2,1)^T$

(d) $(-1,5,6,2)^T$
$(3,-1,2,2)^T$
$(3,0,5,1)^T$
$(3,1,2,0)^T$

2. Show that the given lists of vectors over $\mathbb{C}$ are dependent or independent.

(a) $(-6+2i, -15-9i, 4i, 8-2i)^T$
$(-2-6i, 9-6i, -4, 2+10i)^T$
$(3+4i, 0, 2, -6i)^T$
$(-1-3i, -9+9i, -2, -2+8i)^T$

(b) $(4-6i, 3-6i, 3, 3-6i)^T$
$(8-2i, 1, 3+2i, 1)^T$
$(6+4i, 8-i, 3i, 8-i)^T$
$(4+4i, i, 3i, i)^T$

3. Show that the given sets of vectors over $\mathbb{Z}_7$ are dependent or independent.

(a) $(3,6,3,0)^T$
$(6,1,6,5)^T$
$(1,5,1,4)^T$
$(3,1,3,6)^T$

(b) $(1,6,0,3)^T$
$(2,1,3,6)^T$
$(3,4,1,2)^T$
$(2,2,2,0)^T$

4. Show that the given lists of vectors over $\mathbb{Z}_{11}$ are dependent or independent.

(a) $(3,6,3,0)^T$
$(6,1,6,5)^T$
$(1,5,1,4)^T$
$(3,1,3,6)^T$

(b) $(1,6,0,3)^T$
$(2,1,3,6)^T$
$(3,4,1,2)^T$
$(2,2,2,0)^T$

5. Determine if the following systems of equations have solutions.

(a)
$$\begin{aligned} x - y &= 1 \\ -2x + 2y &= 2 \end{aligned}$$

(b)
$$\begin{aligned} x - 2y &= 3 \\ 3x + y &= 1 \end{aligned}$$

(c)
$$\begin{aligned} x - 2y + z &= 2 \\ -2x - 2y - 3z &= 4 \\ 3x + y + 5z &= 0 \end{aligned}$$

(d)
$$\begin{aligned} x + y + z &= 0 \\ 5x - 2y - 3z &= 4 \\ 7x + 2z &= 0 \end{aligned}$$

(a)
$$A = \begin{pmatrix} 3 & 2 \\ 1 & 5 \end{pmatrix}$$

(b)
$$A = \begin{pmatrix} -1 & 1 & 2 \\ 3 & 1 & 4 \\ -9 & 5 & 8 \end{pmatrix}$$

(c)
$$A = \begin{pmatrix} 1 & -1 & 1 \\ -2 & 2 & 2 \end{pmatrix}$$

(d)
$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 5 & -2 & -3 & 4 \\ 7 & 0 & -1 & 0 \end{pmatrix}$$

6. Determine all X such that $AX = \mathbf{0}$ and X is a column matrix. Find a basis for the row space and give the rank of each matrix A.

(a)
$$A = \begin{pmatrix} 3 & 5 \\ -5 & 3 \end{pmatrix}$$

(b)
$$A = \begin{pmatrix} 2 & 6 & 0 \\ 1 & 3 & 2 \\ -3 & 3 & -2 \end{pmatrix}$$

(c)
$$A = \begin{pmatrix} 3 & 2 & 1 & 5 & -6 \end{pmatrix}$$

(d)
$$A = \begin{pmatrix} -1 & 1 & 2 & 2 \\ 1 & 1 & 0 & 0 \\ 2 & 0 & 4 & 0 \end{pmatrix}$$

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

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

5.7 Coordinate Vectors and Change of Basis

Coordinates are often used to give directions. When we use a printed map, for instance, we transform the units and directions given on the map to match our physical surroundings. One centimeter on the map may correspond to one kilometer of a city and north on a map does not often physically match up with where we are standing (unless we rotate the map, of course).

If we had a smaller friend who is blind-folded, standing in a field some distance from an apple tree, we can give directions to them on how to find the tree. We know which is north and south, but this information would be useless to them. We cannot give them directions on how to move based on compass points and the length of our strides. We cannot say *Take five steps north-east*, for instance. Those would be our coordinates. Knowing the length of their stride to be half the length of one of our strides, we can transform our coordinates into coordinates which they can follow. For example, if we noticed that they are facing north-east, and the apple tree is directly north of them, we can tell them to turn left (their left) 45° and then walk so many paces (their paces, double the number of paces we would need). This is a very simple analogy of what controllers of spacecraft and other remote vehicles do.

We will endeavor to show how to specify coordinates in a mathematical way using basis vectors. We will then show how to give these coordinates in terms of a different basis. In a later section on applications, we will use this method for solving more serious problems than how to get a blind-folded friend to an apple tree.

If the ordered set of vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ form a basis for a vector space $\mathcal{V}$ over a field $\mathbb{F}$, then any vector $\vec{w}$ in $\mathcal{V}$ can be written as a linear combination of these basis vectors

$$\vec{w} = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_n \vec{v}_n$$

for exactly one set of scalars $c_1, c_2, \ldots, c_n$. The scalar $c_i \in \mathbb{F}$ is called the ***i*-th coordinate** of $\vec{w}$ with respect to the given, ordered basis.[4]

It is easy to see why the set of scalar coefficients is unique, because if we were able to define $\vec{w}$ with a different set of scalars, say,

$$\vec{w} = d_1 \vec{v}_1 + d_2 \vec{v}_2 + \cdots + d_n \vec{v}_n$$

subtraction would give

$$\vec{0} = \vec{w} - \vec{w} = (c_1 - d_1)\vec{v}_1 + (c_2 - d_2)\vec{v}_2 + \cdots + (c_n - d_n)\vec{v}_n$$

Since basis vectors are linearly independent, all the coefficients, $(c_i - d_i)$, must equal zero.

At this point we will introduce some notation to denote the coordinate representation of a vector relative to a basis. Note that we list a collection of column vectors in curly braces. This collection is a set and we give it the name E.

[4] Coordinate scalars, c_i, are also called **components**.

Definition 5.22. *Given $E = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_m\}$ is an ordered basis of a vector space $\mathcal{V}$ and given $\vec{w}$ such that*

$$\vec{w} = c_1\vec{e}_1 + c_2\vec{e}_2 + \cdots + c_m\vec{e}_m$$

the ***coordinate vector*** *(or* ***coordinate matrix****) of $\vec{w}$ relative to E is the column matrix in $\mathbb{R}^M$ with entries that are the coordinates of $\vec{w}$.*

$$[\vec{w}]_E = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$$

From the above, we can confidently say that if

$$[\vec{w}]_E = [\vec{x}]_E = (a_1, a_2, \ldots, a_m)^T$$

then

$$\vec{w} = a_1\vec{e}_1 + a_2\vec{e}_2 + \cdots + a_m\vec{e}_m = \vec{x}$$

As an example of a coordinate vector we will use $B = \{(1,3)^T, (4,2)^T\}$ as an ordered basis for $\mathbb{R}^2$. If we have $\vec{v} = (3,4)^T$, what is $[\vec{v}]_B$? We need to solve

$$c_1\begin{pmatrix} 1 \\ 3 \end{pmatrix} + c_2\begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

$$\left(\begin{array}{cc|c} 1 & 4 & 3 \\ 3 & 2 & 4 \end{array}\right) \text{row reduces to} \left(\begin{array}{cc|c} 1 & 0 & \frac{1}{2} \\ 0 & 1 & 1 \end{array}\right)$$

Therefore $c_1 = \frac{1}{2}$ and $c_2 = 1$ so

$$[\vec{v}]_B = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$$

Since this is an example in $\mathbb{R}^2$, we can look at this geometrically.

We mentioned before that vectors $\vec{e}_i$ with all zero entries except the i^{th} entry $= 1$ are standard unit vectors. They are perhaps better known as **standard basis vectors**.

Now we graph the basis B in reference to our standard basis and we now can locate $[\vec{v}]_B = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix}$ as 1/2 along the $\vec{b}_1$ direction and 1 along the $\vec{b}_2$ direction. This is shown in Figure 5.24.

For the general case of $\mathbb{F}^n$, finding a basis is akin to solving a system of n equations in n unknowns. Basis vectors are linearly independent so we may solve the system by finding a matrix inverse.

For some vector space V, if we are given two different ordered bases $U = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ and $W = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ and given some $\vec{v} \in V$ we can find $[\vec{v}]_U$ and $[\vec{v}]_W$. One way is to follow the procedure given in our example above and solve a linear system of n equations for each basis. Then we would have to repeat this for every different $\vec{v}$ we want to convert.

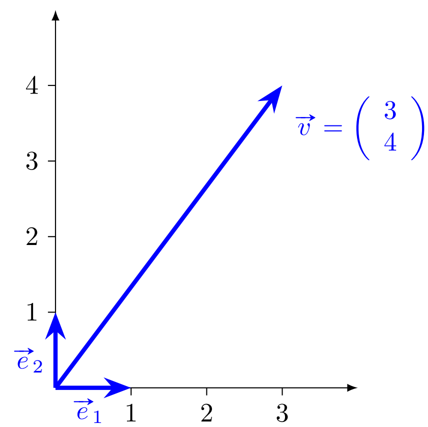


Figure 5.23: $\vec{v}$ located on the standard basis for $\mathbb{R}^2$.

5.8 Affine and Barycentric Coordinates

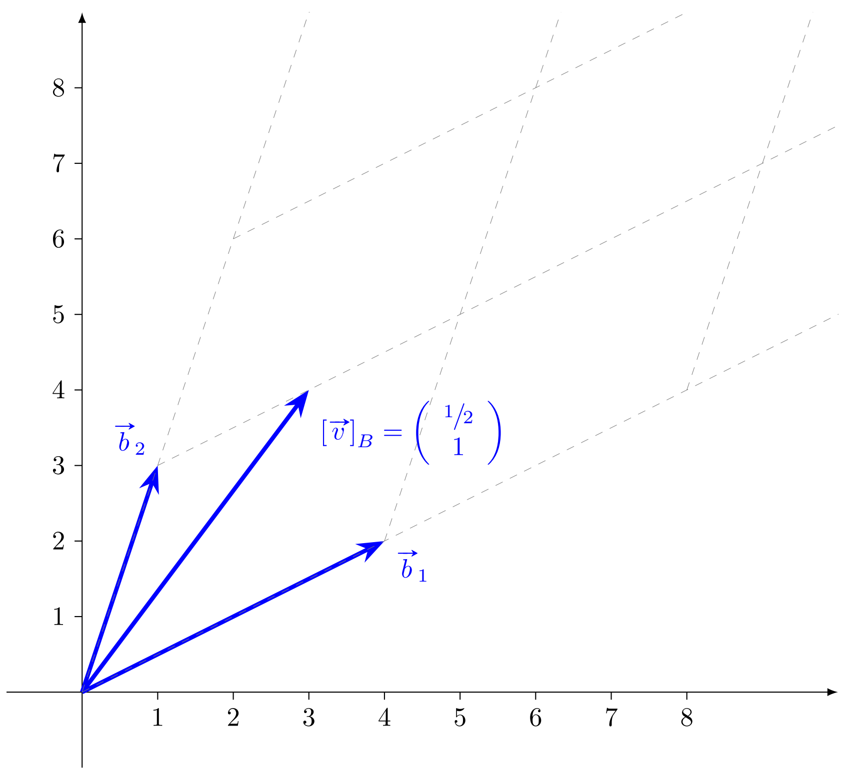


Figure 5.24: $[\vec{v}]_B$ with the B-coordinates superimposed on a standard basis.

Chapter 6

Linear Transformations

We will now see how linear algebra is used with and in other areas of mathematics. We start with a presentation of some concepts using set theory concepts and notation that are standard to the introduction of calculus. We will also make clear what a function is and when a function is linear.

The presentation will not be rigorous as the purpose of the remaining sections of this text is to familiarize the student with aspects of linear algebra. The student will get glimpses of where further study can lead and what concepts are necessary for further understanding. In particular, the vocabulary of more advanced areas of mathematics will be introduced. The aim is to help the student develop a bit of intuition and solid recollection of definitions of important, foundational concepts.

6.1 Definitions and Examples

A relation between two sets X and Y is a set of ordered pairs, each of the form (x, y), where x is a member of X and y is a member of Y. A function from X to Y is a relation between X and Y that has the property that any two ordered pairs with the same x-value also have the same y-value. The variable x is the independent variable, and the variable y is the dependent variable.

Many real-life situations can be modeled by functions. For instance, the area A of a circle is a function of the circle's radius r.

$$A = \pi r^2 \text{ — } A \text{ is a function of } r$$

In this case, r is the independent variable and A is the dependent variable.

Definition of a Real-Valued Function of a Real Variable

Let X and Y be sets of real numbers. A real-valued function f of a real variable x from X to Y is a correspondence that assigns to each number x in X exactly one number y in Y.

The domain of f is the set X. The number y is the image of x under f and is denoted by $f(x)$, which is called the value of f at x. The range of f is a subset of Y and consists of all images of numbers in X. (See Figure.)

Functions can be specified in a variety of ways. In this text, however, you will concentrate primarily on functions that are given by equations involving the dependent and independent variables. For instance, the equation

$$x^2 + 2y = 1 \quad \text{— Equation in implicit form}$$

defines y, the dependent variable, as a function of x, the independent variable. To evaluate this function (that is, to find the y-value that corresponds to a given x-value), it is convenient to isolate y on the left side of the equation.

$$y = \frac{1}{2}(1 - x^2) \quad \text{— Equation in explicit form}$$

Using f as the name of the function, you can write this equation as

$$f(x) = \frac{1}{2}(1 - x^2) \quad \text{— Function notation}$$

The original equation

$$x^2 + 2y = 1$$

implicitly defines y as a function of x. When you solve the equation for y, you are writing the equation in explicit form.

Function notation has the advantage of clearly identifying the dependent variable as $f(x)$ while at the same time telling you that x is the independent variable and that the function itself is 'f.' The symbol $f(x)$ is read 'f of x.'

A function $f : R- > R$ is called '*linear*' if it's of the form $f(x) = ax$. Functions of the form $f(x) = ax + b$ are called '*affine*'.

6.2 Range and Kernel

6.3 Matrix Representation

6.4 Isomorphisms

6.5 Plane Linear Transformations

Chapter 7

Determinants

It is very convenient at times to represent a complex situation with a single number. We do this when we talk about a baseball player's batting average, which gives us no information on how long he has played or even how many base hits he will get on any particular day. Mean temperature, average salary, any average has this property. When a student gets a 90% on a test, this shows that she has a good grasp of the material, but not what particular problems tripped her up. We do find these numbers useful, though, as they can give us definite information. A baseball player with a batting average of .291 has definitely been to first base at least once and the student with 90% certainly knows a good deal of the course material. In order to calculate these numbers, we need to know the entire playing history of the baseball player and we have to know exactly what mark the student got on each question of the test.

The determinant of a matrix is a number that behaves much in this way. It is fussy to calculate but it gives us some very definite information about the matrix.

7.1 Definition of a Determinant

7.2 Geometric Interpretation of a Determinant

7.3 Determinants and Row Operations

7.4 Determinants and Inverses

7.5 Cramer's Rule

Chapter 8

Polar Coordinates and Complex Numbers

The functionality and beauty of complex numbers may only be fully appreciated after one has mastered some calculus. In this chapter we will introduce concepts without full rigor in order to introduce a few applications of complex numbers with regards to linear algebra.

8.1 The Need for Complex Numbers

We will start by re-examining the basic algebraic operations that led to the development of the complex numbers.

An arbitrary linear equation

$$ax + b = 0 \quad (a \neq 0)$$

always has a solution in real numbers as $x = -b/a$.

The following **quadratic equation**[1]

$$x^2 + 1 = 0$$

has no solutions in $\mathbb{R}$. This is because the square of any real number cannot be negative. We have

$$x^2 + 1 \geq 1 > 0$$

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

In order to have solutions to such quadratic equations, the number $i = \sqrt{-1}$ was introduced. That is, $i^2 = -1$. So we may write

$$(x + i)(x - i) = x^2 - ix + ix - (-1) = x^2 + 1$$

[1] a quadratic is a single variable polynomial of degree 2: $ax^2 + bx + c = d$

The root sign, $\sqrt{\ }$, is initially defined to operate with positive, real numbers, so that

$$\sqrt{x}\sqrt{y} = \sqrt{xy}$$

and

$$\sqrt{x}\sqrt{x} = \sqrt{x^2} = x$$

Normally we learn to evaluate roots of positive integer numbers by rewriting them, if possible, as a product of its largest perfect square factor with a non-square factor. For example

$$k = a^2 b$$

Then

$$\sqrt{k} = \sqrt{a^2 b} = \sqrt{a^2}\sqrt{b} = a\sqrt{b}$$

Here, $\sqrt{b}$ is called a **surd**. A surd is an irrational number that is a root of some positive, real number.

8.2 The Complex Plane

8.3 Rotations of the Plane

8.4 Representing Complex Numbers as Matrices

Show that the triangles with the vertices v_1, v_2, v_3 and w_1, w_2, w_3 are similar iff the complex determinant

$$\det \begin{pmatrix} 1 & 1 & 1 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$$

$$\begin{vmatrix} 1 & 1 & 1 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

is zero.

8.5 Other Geometric Vector Representations

Thinking of a vector as a directed, straight line segment is completely suitable for many needs. We may ask, however, are there other ways to draw a vector? After all, there are other forms of geometry other than Euclidean geometry.

8.6 Euler's and de Moivre's Formulas

When we introduced the axioms of a field,we had an exercise where we showed that multiplying two negative numbers gave a positive number.

$$(-x)(-y) = (-x)(-y) + (x)(-y + y)$$
$$= (-x)(-y) + (x)(-y) + (x)(y)$$
$$= (-x + x)(-y) + (x)(y)$$
$$= xy$$

THIS SECTION NEEDS TO BE FIXED:

Item 8 of Definition 1.3 (the definition of a number field) refers to the distributive law; $x * (y + z) = x * y + x * z$ and $(x + y) * z = x * z + y * z$. In this text, for a technical reason that will be discussed, we will alter the distributive law so that we include the condition that, for non-zero scalars under a root sign, $y \neq -z$ and $x \neq -y$. Therefore, we do not allow the product of two negative numbers under a root sign to be evaluated as a positive number, as shown below.

For $a, b > 0$,

$$\sqrt{(-a)(-b)} \neq \sqrt{ab}$$

But

$$\sqrt{(-a)(b)} = \sqrt{(a)(-b)}$$
$$\sqrt{-a} = \sqrt{(-1)(a)} = \sqrt{-1}\sqrt{a}$$

This may seem peculiar at first because this allows the evaluation of a root of a negative number as a negative number, as well as purely imaginary numbers.

Examples;

$$\sqrt{(-5)^2} = \sqrt{(-5)(-5)} = \sqrt{(-1)(-1)(5)(5)} = \sqrt{(-1)}\sqrt{(-1)}\sqrt{25} = i^2 5 = -5$$
$$\sqrt{(5)(-5)} = \sqrt{25(-1)} = 5i$$
$$\sqrt{(-5)(-5)(-1)} = \sqrt{(-5)(-5)}\sqrt{(-1)} = -5i$$

Defining multiplication of two negative real numbers under a root sign this way is known in complex analysis as a **branch cut**. We have chosen to do it this way solely to give a simple algebraic reason to discount the following "proof";

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = i^2 = -1$$

This is the technical reason we alluded to previously. We will not pursue this matter further in this text.

We will show that complex arithmetic has a geometric interpretation by introducing the complex plane. The complex plane is used to represent the real coefficients of a complex number, $x + yi$, by using x and y as the coordinates on the x-axis and y-axis of a standard rectangular coordinate system. Positions may also be determined by two unique numbers based on the vector which represents this point. Specifically, we identify a point by the associated vector's magnitude (distance from the origin) and the angle the vector makes with the x-axis. These two numbers are the **polar coordinates**. On a plane, polar coordinates are like rectangular coordinates and they are ordered pairs of numbers assigned to each point on the plane.

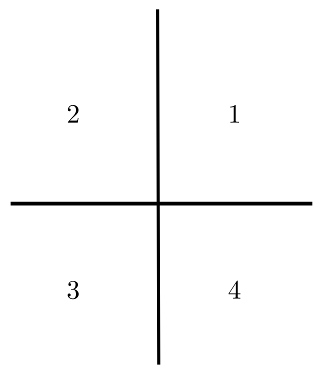


Figure 8.1: Quadrants of a plane

We would like to convert between the rectangular x and y coordinates and the polar coordinates comprised of a magnitude and an angle. In order to do so, we must ensure that the polar coordinates are unique, so that we may go back and forth between the two coordinate systems without getting different answers. We do this by only allowing angles less than 2π radians and also by restricting the angles one may use for each trigonometric function to specific quadrants on the plane. Figure 8.1 shows the four standard quadrants. We will begin by concentrating on the first and third quadrants and we will set limits on what radian angle measures we will allow for the three most fundamental trigonometric functions: sine, cosine and tangent.

Chapter 9

The Geometric Product

During our travels through the mysteries of linear algebra, we have come across many ways of defining a product. We have used the product of two rationals, the product of two reals, two complex numbers, two matrices, a scalar with a vector and a scalar with a matrix. We have also defined and used an inner product of two vectors, the standard dot product, which returns a scalar. As well, we have been introduced to an outer product of two orthonormal vectors, which returns a directed area, and the related outer product of three orthonormal vectors, which returns a directed volume.

It is probably not obvious, but what we really would like to have is another product which is called the **geometric product**. It is an extension of the products we have used before. The geometric product allows one to multiply multivectors, which are objects made up of linear combinations of **blades**.

9.1 Blades

Definition 9.1. *A **bound vector** is an element of a **bound vector space**. A bound vector space contains, besides free vectors, one or more other elements called points. A bound vector space always has one special point, $*\mathbf{0}$, called the **origin**.*

Geometrically we usually think of a bound vector in $\mathbb{R}^2$ as a point on a plane defining a directed line segment from the *origin* of a chosen axis. Figure 5.1 shows three examples of bound vectors in $\mathbb{R}^2$. Note that there is no problem with thinking of a bound vector as indistinguishable from a point, as long as there is exactly one vector for every point in the space and the origin is indistinguishable from the zero vector ($*\mathbf{0} = \vec{0}$). In other words, given an axis with an origin and directed line segments all emanating from that origin, these directed line segments form a vector space and there is no difference between a point and a vector. The importance of distinguishing between vectors and points will not become apparent until we introduce a type of vector space that contains points and *multivectors*, a space where *adding two points together is different from* *adding two vectors together*. We will not do this until after we discuss other ways of augmenting a vector space.

It might be obvious by now that not all directed line segments are bound vectors, unless they have the same starting point, the origin. Later on we will use vector subtraction to show how to translate any directed line segment into a bound vector (which *does* emanate from the origin).

Our purpose in adopting this notation is to give a geometric interpretation to many of the ideas which we will encounter in the rest of this text (and to save some typing). We will construct geometric objects in $\mathbb{R}^3$ and show how to represent these objects using the definitions and theorems of linear algebra. We will manipulate these objects using matrix arithmetic in order to shrink them, expand them, rotate them and alter them in other ways. We will then use the methods we develop to apply to the solution of problems in other areas. We start with a discussion of the use of unit vectors and unit bivectors on a plane and then extend these concepts to $\mathbb{R}^3$.

9.2 Multivectors

Multivectors are, among other things, a very flexible way of modelling geometrical objects.

Chapter 10

More Applications

We come, sadly, to the end our little introduction to linear algebra. Although a vast world of applications await those who undertake a study of calculus, set theory and more advanced linear algebra, the exercises in this chapter should reinforce the student's confidence in the usefulness of her or his hard-won knowledge.

10.1 Derivation of $\sum k^n$

Occasionally we will find that we must evaluate equations like $y = 1^2+2^2+3^2+4^2$ or $x = 1^n + 2^n + 3^n + \cdots + k^n$. It would be useful to have functions which would give us exact answers for arbitrary values of k and n without having to perform expansions and huge amounts of arithmetic. We will show a method for finding such functions in this section.

For our purposes we define a **sequence** as an ordered list of scalars. In other words, there is a first scalar, a second, and so on. An **infinite sequence** is one that continues forever. Sequences are often represented by an algorithm which generates the next term in the sequence.

A **series** is the sum of the members of a sequence. An **infinite series** is the sum of an infinite sequence. We can write an infinite series using the symbols $\sum^{\infty}$. If the members of a sequence tend to become smaller as we continue down the list, then sometimes the infinite series made from an infinite sequence of this kind **converges** to a value. Calculus is needed to understand infinite series. We are only interested in finite series in this text, all of which obviously can be evaluated as a finite number.

Given a positive integer k we define $k! = k * (k-1) * (k-2) * \cdots * 3 * 2 * 1$. We say $k!$ is k **factorial**. Thus, $1! = 1$, $2! = 2 * 1 = 2$, $3! = 3 * 2 * 1 = 6$, $4! = 4 * 3 * 2 * 1 = 24$, and so on. We also define $0! = 1$. Factorials are used to calculate the number of ways objects may be arranged in order. For the letters a, b, c there are $3! = 6$ ways to arrange them. They are abc, acb, bac, bca, cab and cba. Each arrangement is called a **permutation**.

If we have 4 letters, a, b, c and d, the number of ways we can arrange them is $4! = 24$. How many ways are there to take the four letters two at a time? We supply the list below.

$$\begin{array}{cccc} ab & ba & ca & da \\ ac & bc & cb & db \\ ad & bd & cd & dc \end{array}$$

The formula to calculate the number of k distinct objects taken r at a time is

$$\frac{k!}{(k-r)!}$$

This formula holds when $r \leq k$ and is defined to be zero when $r > k$.

In the list of permutations we gave, notice that we have ab and ba counted separately. This is because order is important. If order is not important, and we wish to know how many **combinations** we can have, then the formula is

$$\binom{k}{r} = \frac{k!}{r!(k-r)!}$$

Please note that the use of braces in the left part of the equation *does not specify a matrix.* In this context it stands for the **choose function**. It means we require the number of combinations of k distinct objects chosen r at a time. We read it as k *choose* r. A table listing the combinations of choices of two letters from the four letters *abcd* is

$$\begin{array}{lll} ab & & \\ ac & bc & \\ ad & bd & cd \end{array}$$

The number of combinations is given by

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4*3*2*1}{2*1*2*1} = 6$$

We will be interested in acquiring the coefficients when we expand $(x+y)^n$ for $x, y \in \mathbb{R}$ and some positive integer n. We will use the choose function to state the **binomial theorem**.

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^j y^{n-j}$$

We will not prove this theorem here. We are interested in the case when $x = i$ and $y = 1$ for some $i \in \mathbb{Z}^+$ (i is a positive integer). The theorem then gives us

$$(i+1)^n = \sum_{j=0}^{n} \binom{n}{j} i^j$$

$$\text{But } \binom{n}{0} i^0 = 1 \text{ and } \binom{n}{n} i^n = i^n$$

So we can simplify and write

$$(i+1)^n = 1 + \sum_{j=1}^{n-1} \binom{n}{j} i^j + i^n$$

$$(i+1)^n - i^n = 1 + \sum_{j=1}^{n-1} \binom{n}{j} i^j$$

We can sum both sides of the equation with i from 1 to m.

$$\sum_{i=1}^{m} \Big((i+1)^n - i^n \Big) = \sum_{i=1}^{m} \Big(1 + \sum_{j=1}^{n-1} \binom{n}{j} i^j \Big)$$

Because we are adding and subtracting many of the same quantities, the left side of the equation can be simplified to

$$\sum_{i=1}^{m} \Big((i+1)^n - i^n \Big) = (m+1)^n - 1$$

The right side of the equation can also be rewritten as

$$\sum_{i=1}^{m} \Big(1 + \sum_{j=1}^{n-1} \binom{n}{j} i^j \Big) = \sum_{i=1}^{m} 1 + \sum_{i=1}^{m} \sum_{j=1}^{n-1} \binom{n}{j} i^j = m + \sum_{j=1}^{n-1} \binom{n}{j} \sum_{i=1}^{m} i^j$$

Replacing the simplified sections back into the original equation gives us

$$(m+1)^n - 1 = m + \sum_{j=1}^{n-1} \binom{n}{j} \sum_{i=1}^{m} i^j$$

If we let $m = k - 1$, then we have

$$k^n = k + \sum_{j=1}^{n-1} \binom{n}{j} \sum_{i=1}^{k-1} i^j$$

If we define a function $S_p(k)$ as

$$S_p(k) = \sum_{i=1}^{k} i^p$$

we can then, finally, write the equation we would like to use

$$k^n = k + \sum_{j=1}^{n-1} \binom{n}{j} S_j(k-1)$$

Let us use the above formula to write out a system for the first few equations.

$$
\begin{array}{llcl}
n & & & \\
1 & k & = & k \\
2 & k^2 & = & k + 2S_1(k-1) \\
3 & k^3 & = & k + 3S_1(k-1) + 3S_2(k-1) \\
4 & k^4 & = & k + 4S_1(k-1) + 6S_2(k-1) + 4S_3(k-1) \\
5 & k^5 & = & k + 5S_1(k-1) + 10S_2(k-1) + 10S_3(k-1) + 5S_4(k-1) \\
\vdots & & &
\end{array}
$$

For $n = 4$, we can write the matrix equation for the above system as

$$
\begin{pmatrix} k \\ k^2 \\ k^3 \\ k^4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 3 & 0 \\ 1 & 4 & 6 & 4 \end{pmatrix} \begin{pmatrix} k \\ S_1(k-1) \\ S_2(k-1) \\ S_3(k-1) \end{pmatrix}
$$

The 4×4 matrix is triangular and all the diagonal entries are non-zero so we can compute the inverse and write

$$
\begin{pmatrix} k \\ S_1(k-1) \\ S_2(k-1) \\ S_3(k-1) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{6} & -\frac{1}{2} & \frac{1}{3} & 0 \\ 0 & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} k \\ k^2 \\ k^3 \\ k^4 \end{pmatrix}
$$

$S_i(k-1)$ gives us sums of $(k-1)$ terms. We'd like k terms so we will use the fact that $S_n(k) = S_n(k-1) + k^n$. If we call the $n \times n$ matrix A, then this means we will add 1 to $a_{i+1,i}$ (the number immediately below the main diagonal scalars).

$$
\begin{pmatrix} k \\ S_1(k) \\ S_2(k) \\ S_3(k) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} k \\ k^2 \\ k^3 \\ k^4 \end{pmatrix}
$$

From our definition $S_2(3) = \sum_{i=1}^{3} i^2 = 1^2 + 2^2 + 3^2$. By matrix multiplication we find that $S_2(k) = \frac{1}{6}k + \frac{1}{2}k^2 + \frac{1}{3}k^3$. So then $S_2(3) = \frac{1}{6} * 3 + \frac{1}{2} * 3^2 + \frac{1}{3} * 3^3 = 14$. Obviously for large values of k there is significantly less arithmetic involved in computing the sum using the formulas generated by the matrix inverse.

The numbers in the first column of the final form of $A_{n \times n}$ are the first n Bernoulli numbers. These numbers appear in many applications. For example, using calculus, these numbers are used to generate an infinite series which represents the cotangent function. Euler applied these numbers to calculate exact values for infinite series of the form

$$\sum_{k=1}^{\infty} \frac{1}{k^{2n}}$$

No one has yet to show exact values for

$$\sum_{k=1}^{\infty} \frac{1}{k^{n}}$$

when n is odd and greater than 1.

Exercises 10.1

1. Calculate $S_4(k)$ and use the resulting equation to evaluate $\sum_{k=1}^{10} k^4$.

10.2 Area and Volume Calculations

10.3 Statistics and Least Squares Regression

Much of applied science, statistics, engineering and other areas use mathematics to find simple equations which approximately model observations. The most basic of these kind of models is to fit a straight line to a bunch of points plotted on a grid. We would like an equation $y = mx + b$ which gives us y values for any given x value with fairly good accuracy. But what do we mean by fairly good accuracy? How do we define a straight line to have the best fit? In other words, we need values for m and b in our linear equation which gives a line which passes as close as possible to as many points as possible. Once we examine the problem in its particulars, it will not prove as daunting as it may appear now.

10.4 Quadratic Approximation

10.5 Linear Programming

10.6 3D Computer Graphics

L'Envoi

There are many[1] very good books on linear algebra available that have been written for university students. I cannot, therefore, suggest any particular one as tastes and interests of students and instructors differ dramatically. I have, however, listed a few that I found exemplary in the References. I do suggest that a complete course in calculus be undertaken before preceding with further studies in linear algebra.

Geometric algebra is still a tiny bit controversial as many people who have mastered other methods do not see a pressing need to add further complexities to their areas. I look upon the use of geometric algebra much in the same way as I view the use of a mouse with a computer. The adoption of the mouse (and touchscreen) does add extra hardware and the need for more complex software, but once one starts using a mouse (or touchscreen) one never wants to go back to working without it. I feel the same way about LaTeX, which was used to write this text. Anyone who needs to create mathematical manuscripts (or any type of beautiful document, actually) is advised to spend a few days introducing themselves to LaTeX. After more than 40 years of use by professionals, it is still the best.

Lampp was a work of love and I must say that I wish I had it to play with when I was learning linear algebra, which was the whole reason for this book.

[1] many, many, many

Answers to Selected Problems

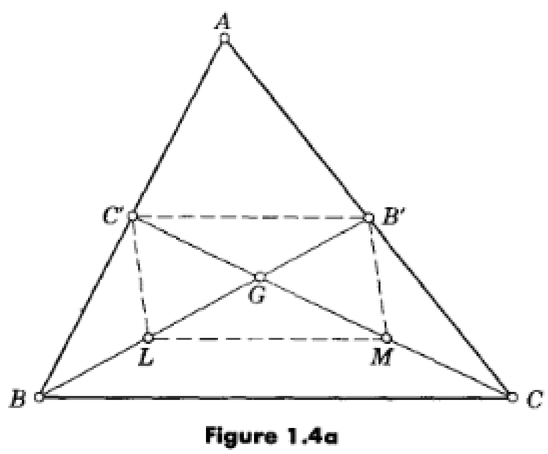


Figure 10.1: Figure 1.4a

From *Introduction to Geometry, second edition* by H. S. M. Coxeter, F. R. S., pg. 10-11

VI.2. If a straight line be drawn parallel to one side of a triangle, it will cut the other sides proportionately; and, if two sides of the triangle be cut proportionately, the line joining the points of section will be parallel to the remaining side.

VI.4. If corresponding angles of two triangles are equal, then corresponding sides are proportional.

The line joining a vertex of a triangle to the midpoint of the opposite side is called a median.

Let two of the three medians, say BB' and CC, meet in G (Figure 1.4a).

Let L and M be the midpoints of GB and GC. By Euclid VI.2 and 4 (which were quoted on page 8), both C'B' and LM are parallel to BC and half as long. Therefore B'CLM is a parallelogram. Since the diagonals of a parallelogram bisect each other, we have

B'G = GL = LB, C'G = GM = MC

Thus the two medians BB', CC trisect each other at G. In other words, this point G, which could have been denned as a point of trisection of one median, is also a point of trisection of another, and similarly of the third. We have thus proved [by the method of Court 1, p. 58] the following theorem:

1.41 The three medians of any triangle all pass through one point. This common point G of the three medians is called the centroid of the triangle. Archimedes (c. 287-212 b.c.) obtained it as the center of gravity of a triangular plate of uniform density.

Prove

1. Any triangle having two equal medians is isosceles.

2. The sum of the medians of a triangle lies between $\frac{3}{4}$ p and p, where p is the sum of the sides. [Court I, pp. 60-61.]

From Stillwell (without proof):

Prove that the centroid of triangle formed from the endpoints of 3 vectors, $\vec{a}$, $\vec{b}$ and $\vec{c}$, is located at $\frac{1}{3}(\vec{a} + \vec{b} + \vec{c})$

VARIGNON'S PARALLELOGRAM

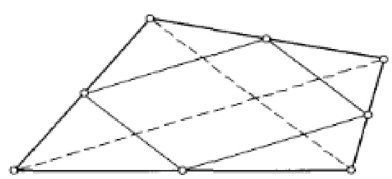


Figure 13.2g

Figure 10.2: 13.2g

The midpoints of the four sides of any simple quadrangle are the vertices of a parallelogram (Figure 13.2g; cf. Figure 4.2c). This theorem was discovered by Pierre Varignon (1654-1722). It shows that the bimedians, which join the midpoints of opposite sides of the quadrangle, bisect each other. Thus the corollary to Hjelmslev's theorem (§ 3.6) becomes an affine theorem when we replace the hypotheses 3.61 by $AB = BC, A'B' = B'C$.

Fermat's little theorem - pascal's triangle - Ellenberg talks about rotating n-dimensional cube

Answers to Selected Exercises

Appendix A

Lampp on your Computer

Lampp, the Linear Algebra Matrix aPP (named to coincide with Linear Algebra for Mouse, Pen and Pad) is an application program written in JavaFX. It was designed to be cross-platform and work on many devices. It can be downloaded from the Google Play Store for Android devices. It can also be downloaded for an iPad device from the iTunes app store. The Windows version may be downloaded from the Windows App Store.

Linux users or Windows users can obtain the program from the included CD. For Windows or Linux users, one must first install Java 1.8 or higher on one's PC or tablet. Then one can copy the Lampp.jar file to one's PC.

Links to the latest versions are also available from the textbook website.

References

[1] Tom M. Apostol and Mamikon A. Mnatsakanian. *New Horizons in Geometry*. The Mathematical Association of America, 2012.

[2] Sheldon Jay Axler. *Linear Algebra Done Right*. Springer, 2015.

[3] Alexander Basilevsky. *Applied Matrix Algebra in the Statistical Sciences*. Dover, 2005.

[4] David Carlson, Charles R. Johnson, David C. Lay, and A. Duane Porter. The linear algebra curriculum study group recommendations for the first course in linear algebra. *The College Mathematics Journal*, 42:41–46, 1993.

[5] G. Chrystal. *Algebra : An Elementary Text-Book, Part I & II*. Dover, 1904.

[6] Charles G. Cullen. *Matrices and Linear Transformations*. Dover, 1990.

[7] Desmond Fearnley-Sanders. Hermann grassmann and the creation of linear algebra. *American Mathematical Monthly*, 86:809–817, 1979.

[8] Paul A. Fuhrmann. *A Polynomial Approach to Linear Algebra*. Springer, 2012.

[9] Robert Gauntt, Gustave Rabson, Carl Cohen, and J. P. Ballantine. Classroom notes. *The American Mathematical Monthly*, 63(4):247–252, 1956.

[10] Guershon Harel. The linear algebra curriculum study group recommendations: Moving beyond concept definition. *Resources for Teaching Linear Algebra,. MAA Notes*, 24:107–126, 1997.

[11] David Hestenes. Oersted medal lecture 2002: Reforming the mathematical language of physics. `http://geocalc.clas.asu.edu/html/Oersted-ReformingTheLanguage.html`. Last Accessed: 2016-03-16.

[12] David Hestenes. Primer on geometric algebra for introductory mathematics and physics. `http://geocalc.clas.asu.edu/pdf/PrimerGeometricAlgebra.pdf`. Last Accessed: 2016-03-16.

[13] David C. Lay. *Linear Algebra and Its Applications*. Addison-Wesley, 2012.

[14] Jiri Matousek. *Thirty-three Miniatures: Mathematical and Algorithmic Applications of Linear Algebra.* American Mathematical Society, 2010.

[15] Gregory H. Moore. The axiomatization of linear algebra: 1875-1940. *Historia Mathematica*, 22:162–303, 1995.

[16] David C. Poole. *Linear Algebra: A Modern Introduction.* Cengage Learning, 2015.

[17] H. E. Slaught and N.J. Lennes. *Solid Geometry with Problems and Applications (Revised edition).* Allyn and Bacon, 1919.

[18] Garret Sobczyk. *New Foundations in Mathematics : The Geometric Concept of Number.* Birkhauser, 2013.

[19] Gilbert Strang. *Introduction to Linear Algebra.* Wellesley - Cambridge Press, 2009.

[20] Thomas Yuster. The reduced row echelon form of a matrix is unique: A simple proof. *Mathematics Magazine*, 57:93–94, 1984.