Elementary Linear Algebra

Kuttler

January 10, 2012

Contents

Preface

This is an introduction to linear algebra. The main part of the book features row operations and everything is done in terms of the row reduced echelon form and specific algorithms. At the end, the more abstract notions of vector spaces and linear transformations on vector spaces are presented. However, this is intended to be a first course in linear algebra for students who are sophomores or juniors who have had a course in one variable calculus and a reasonable background in college algebra. I have given complete proofs of all the fundamental ideas, but some topics such as Markov matrices are not complete in this book but receive a plausible introduction. The book contains a complete treatment of determinants and a simple proof of the Cayley Hamilton theorem although these are optional topics. The Jordan form is presented as an appendix. I see this theorem as the beginning of more advanced topics in linear algebra and not really part of a beginning linear algebra course. There are extensions of many of the topics of this book in my on line book [11]. I have also not emphasized that linear algebra can be carried out with any field although there is an optional section on this topic, most of the book being devoted to either the real numbers or the complex numbers. It seems to me this is a reasonable specialization for a first course in linear algebra.

Linear algebra is a wonderful interesting subject. It is a shame when it degenerates into nothing more than a challenge to do the arithmetic correctly. It seems to me that the use of a computer algebra system can be a great help in avoiding this sort of tedium. I don't want to over emphasize the use of technology, which is easy to do if you are not careful, but there are certain standard things which are best done by the computer. Some of these include the row reduced echelon form, PLU factorization, and QR factorization. It is much more fun to let the machine do the tedious calculations than to suffer with them yourself. However, it is not good when the use of the computer algebra system degenerates into simply asking it for the answer without understanding what the oracular software is doing. With this in mind, there are a few interactive links which explain how to use a computer algebra system to accomplish some of these more tedious standard tasks. These are obtained by clicking on the symbol ►. I have included how to do it using maple and scientific notebook because these are the two systems I am familiar with and have on my computer. Other systems could be featured as well. It is expected that people will use such computer algebra systems to do the exercises in this book whenever it would be helpful to do so, rather than wasting huge amounts of time doing computations by hand. However, this is not a book on numerical analysis so no effort is made to consider many important numerical analysis issues.

Some Prerequisite Topics

The reader should be familiar with most of the topics in this chapter. However, it is often the case that set notation is not familiar and so a short discussion of this is included first. Complex numbers are then considered in somewhat more detail. Many of the applications of linear algebra require the use of complex numbers, so this is the reason for this introduction.

1.1 Sets And Set Notation

A set is just a collection of things called elements. Often these are also referred to as points in calculus. For example $\{1,2,3,8\}$ would be a set consisting of the elements 1,2,3, and 8. To indicate that 3 is an element of $\{1,2,3,8\}$, it is customary to write $3 \in \{1,2,3,8\}$. $9 \notin \{1,2,3,8\}$ means 9 is not an element of $\{1,2,3,8\}$. Sometimes a rule specifies a set. For example you could specify a set as all integers larger than 2. This would be written as $S = \{x \in \mathbb{Z} : x > 2\}$. This notation says: the set of all integers, x, such that $x > 2$.

If A and B are sets with the property that every element of A is an element of B, then A is a subset of B. For example, $\{1,2,3,8\}$ is a subset of $\{1,2,3,4,5,8\}$, in symbols, $\{1,2,3,8\} \subseteq \{1,2,3,4,5,8\}$. It is sometimes said that "A is contained in B" or even "B contains A". The same statement about the two sets may also be written as $\{1,2,3,4,5,8\} \supseteq \{1,2,3,8\}$.

The union of two sets is the set consisting of everything which is an element of at least one of the sets, A or B. As an example of the union of two sets $\{1,2,3,8\} \cup \{3,4,7,8\} = \{1,2,3,4,7,8\}$ because these numbers are those which are in at least one of the two sets. In general

$$A \cup B \equiv \{x : x \in A \text{ or } x \in B\}.$$

Be sure you understand that something which is in both A and B is in the union. It is not an exclusive or.

The intersection of two sets, A and B consists of everything which is in both of the sets. Thus $\{1,2,3,8\} \cap \{3,4,7,8\} = \{3,8\}$ because 3 and 8 are those elements the two sets have in common. In general,

$$A \cap B \equiv \{x : x \in A \text{ and } x \in B\}.$$

The symbol $[a,b]$ where a and b are real numbers, denotes the set of real numbers x, such that $a \leq x \leq b$ and $[a,b)$ denotes the set of real numbers such that $a \leq x < b$. (a,b) consists of the set of real numbers x such that $a < x < b$ and $(a,b]$ indicates the set of numbers x such that $a < x \leq b$. $[a,\infty)$ means the set of all numbers x such that $x \geq a$ and $(-\infty,a]$ means the set of all real numbers which are less than or equal to a. These sorts of sets of real numbers are called intervals. The two points a and b are called endpoints of the interval. Other intervals such as $(-\infty,b)$ are defined by analogy to what was just explained. In general, the curved parenthesis indicates the end point it sits next to is not included while the square parenthesis indicates this end point is included. The reason that there will always be a curved parenthesis next to ∞ or $-\infty$ is that these are not real numbers. Therefore, they cannot be included in any set of real numbers.

A special set which needs to be given a name is the empty set also called the null set, denoted by $\emptyset$. Thus $\emptyset$ is defined as the set which has no elements in it. Mathematicians like to say the empty set is a subset of every set. The reason they say this is that if it were not so, there would have to exist a set A, such that $\emptyset$ has something in it which is not in A. However, $\emptyset$ has nothing in it and so the least intellectual discomfort is achieved by saying $\emptyset \subseteq A$.

If A and B are two sets, $A \setminus B$ denotes the set of things which are in A but not in B. Thus

$$A \setminus B \equiv \{x \in A : x \notin B\}.$$

Set notation is used whenever convenient.

To illustrate the use of this notation relative to intervals consider three examples of inequalities. Their solutions will be written in the notation just described.

Example 1.1.1 *Solve the inequality* $2x + 4 \leq x - 8$

$x \leq -12$ is the answer. This is written in terms of an interval as $(-\infty, -12]$.

Example 1.1.2 *Solve the inequality* $(x + 1)(2x - 3) \geq 0$.

The solution is $x \leq -1$ or $x \geq \frac{3}{2}$. In terms of set notation this is denoted by $(-\infty, -1] \cup [\frac{3}{2}, \infty)$.

Example 1.1.3 *Solve the inequality* $x(x + 2) \geq -4$.

This is true for any value of x. It is written as $\mathbb{R}$ or $(-\infty, \infty)$.

1.2 Functions

The concept of a function is that of something which gives a unique output for a given input.

Definition 1.2.1 *Consider two sets, D and R along with a rule which assigns a unique element of R to every element of D. This rule is called a **function** and it is denoted by a letter such as f. Given $x \in D$, $f(x)$ is the name of the thing in R which results from doing f to x. Then D is called the **domain** of f. In order to specify that D pertains to f, the notation $D(f)$ may be used. The set R is sometimes called the **range** of f. These days it is referred to as the **codomain**. The set of all elements of R which are of the form $f(x)$ for some $x \in D$ is therefore, a subset of R. This is sometimes referred to as the image of f. When this set equals R, the function f is said to be **onto**, also **surjective**. If whenever $x \neq y$ it follows $f(x) \neq f(y)$, the function is called **one to one.** , also **injective** It is common notation to write $f : D \mapsto R$ to denote the situation just described in this definition where f is a function defined on a domain D which has values in a codomain R. Sometimes you may also see something like $D \overset{f}{\mapsto} R$ to denote the same thing.*

Example 1.2.2 *Let D consist of the set of people who have lived on the earth except for Adam and for $d \in D$, let $f(d) \equiv$ the biological father of d. Then f is a function.*

This function is not the sort of thing studied in calculus but it is a function just the same.

Example 1.2.3 *Consider the list of numbers $\{1, 2, 3, 4, 5, 6, 7\} \equiv D$. Define a function which assigns an element of D to $R \equiv \{2, 3, 4, 5, 6, 7, 8\}$ by $f(x) \equiv x + 1$ for each $x \in D$.*

This function is onto because every element of R is the result of doing f to something in D. The function is also one to one. This is because if $x + 1 = y + 1$, then it follows $x = y$. Thus different elements of D must go to different elements of R.

In this example there was a clearly defined procedure which determined the function. However, sometimes there is no discernible procedure which yields a particular function.

Example 1.2.4 *Consider the ordered pairs,* $(1,2),(2,-2),(8,3),(7,6)$ *and let*

$$D \equiv \{1,2,8,7\},$$

the set of first entries in the given set of ordered pairs and let

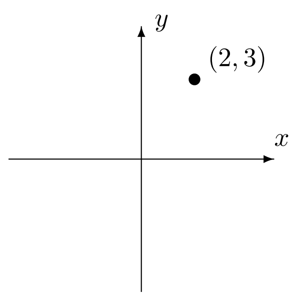
$$R \equiv \{2,-2,3,6\},$$

the set of second entries, and let $f(1)=2, f(2)=-2, f(8)=3,$ *and* $f(7)=6$.

This specifies a function even though it does not come from a convenient formula.

1.3 Graphs Of Functions

Recall the notion of the Cartesian coordinate system you probably saw earlier. It involved an x axis, a y axis, two lines which intersect each other at right angles and one identifies a point by specifying a pair of numbers. For example, the number $(2,3)$ involves going 2 units to the right on the x axis and then 3 units directly up on a line perpendicular to the x axis. For example, consider the following picture.



Because of the simple correspondence between points in the plane and the coordinates of a point in the plane, it is often the case that people are a little sloppy in referring to these things. Thus, it is common to see (x,y) referred to as a point in the plane. In terms of relations, if you graph the points as just described, you will have a way of visualizing the relation.

The reader has likely encountered the notion of graphing relations of the form $y=2x+3$ or $y=x^2+5$. The meaning of such an expression in terms of defining a relation is as follows. The relation determined by the equation $y=2x+3$ means the set of all ordered pairs (x,y) which are **related** by this formula. Thus the relation can be written as

$$\{(x,y): y=2x+3\}.$$

The relation determined by $y=x^2+5$ is

$$\left\{(x,y): y=x^2+5\right\}.$$

Note that these relations are also functions. For the first, you could let $f(x)=2x+3$ and this would tell you a rule which tells what the function does to x. However, some relations are not functions. For example, you could consider $x^2+y^2=1$. Written more formally, the relation it defines is

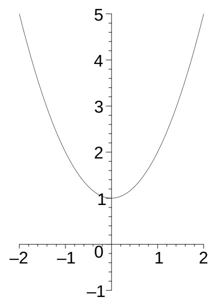
$$\left\{(x,y): x^2+y^2=1\right\}$$

Now if you give a value for x, there might be two values for y which are associated with the given value for x. In fact

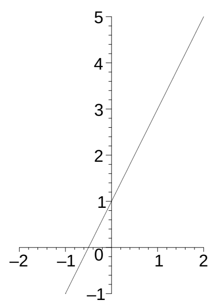
$$y = \sqrt[\pm]{1-x^2}$$

Thus this relation would not be a function.

Recall how to graph a relation. You first found lots of ordered pairs which satisfied the relation. For example $(0,3)$,$(1,5)$, and $(-1,1)$ all satisfy $y = 2x + 3$ which describes a straight line. Then you connected them with a curve. Here are some simple examples which you should see that you understand. First here is the graph of $y = x^2 + 1$.



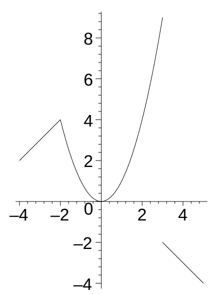
Now here is the graph of the relation $y = 2x + 1$ which is a straight line.



Sometimes a relation is defined using different formulas depending on the location of one of the variables. For example, consider

$$y = \begin{cases} 6 + x & \text{if} \quad x \leq -2 \\ x^2 & \text{if} \quad -2 < x < 3 \\ 1 - x & \text{if} \quad x \geq 3 \end{cases}$$

Then the graph of this relation is sketched below.



A very important type of relation is one of the form $y - y_0 = m(x - x_0)$, where m, x_0, and y_0 are numbers. The reason this is important is that if there are two points, (x_1, y_1), and (x_2, y_2) which satisfy this relation, then

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{(y_1 - y_0) - (y_2 - y_0)}{x_1 - x_2} = \frac{m(x_1 - x_0) - m(x_2 - x_0)}{x_1 - x_2}$$
$$= \frac{m(x_1 - x_2)}{x_1 - x_2} = m.$$

Remember from high school, the slope of the line segment through two points is always the difference in the y values divided by the difference in the x values, taken in the same order. Sometimes this is referred to as the rise divided by the run. This shows that there is a constant slope m, the slope of the line, between any pair of points satisfying this relation. Such a relation is called a straight line. Also, the point (x_0, y_0) satisfies the relation. This is more often called the equation of the straight line.

Geometrically, this means the graph of the relation is a straight line because the slope between any two points is always the same.

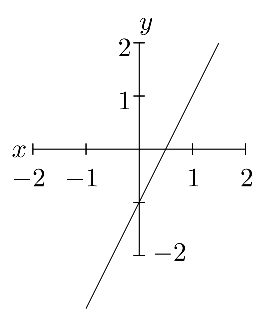
Example 1.3.1 *Find the relation for a straight line which contains the point* $(1, 2)$ *and has constant slope equal to* 3.

From the above discussion, $(y - 2) = 3(x - 1)$.

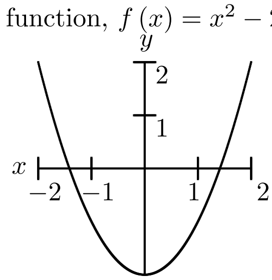
Definition 1.3.2 *Let* $f : D(f) \mapsto R(f)$ *be a function. The graph of* f *consists of the set*

$$\{(x, y) : y = f(x) \text{ for } x \in D(f)\}.$$

Note that knowledge of the graph of a function is equivalent to knowledge of the function. To find $f(x)$, simply observe the ordered pair which has x as its first element and the value of y equals $f(x)$.

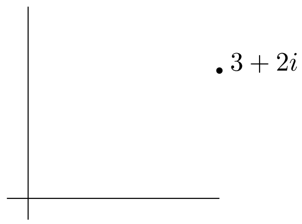


Here is the graph of the function, $f(x) = x^2 - 2$



1.4 The Complex Numbers

Recall that a real number is a point on the real number line. Just as a real number should be considered as a point on the line, a complex number is considered a point in the plane which can be identified in the usual way using the Cartesian coordinates of the point. Thus (a, b) identifies a point whose x coordinate is a and whose y coordinate is b. In dealing with complex numbers, such a point is written as $a + ib$. For example, in the following picture, I have graphed the point $3 + 2i$. You see it corresponds to the point in the plane whose coordinates are $(3, 2)$.



Multiplication and addition are defined in the most obvious way subject to the convention that $i^2 = -1$. Thus,

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

and

$$\begin{aligned}(a + ib)(c + id) &= ac + iad + ibc + i^2 bd \\ &= (ac - bd) + i(bc + ad).\end{aligned}$$

Every non zero complex number $a + ib$, with $a^2 + b^2 \neq 0$, has a unique multiplicative inverse.

$$\frac{1}{a + ib} = \frac{a - ib}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2}.$$

You should prove the following theorem.

Theorem 1.4.1 *The complex numbers with multiplication and addition defined as above form a field satisfying all the field axioms. These are the following list of properties.*

1. $x + y = y + x$, *(commutative law for addition)*
2. $x + 0 = x$, *(additive identity).*
3. *For each* $x \in \mathbb{R}$, *there exists* $-x \in \mathbb{R}$ *such that* $x + (-x) = 0$, *(existence of additive inverse).*
4. $(x + y) + z = x + (y + z)$, *(associative law for addition).*
5. $xy = yx$, *(commutative law for multiplication). You could write this as* $x \times y = y \times x$.
6. $(xy)z = x(yz)$, *(associative law for multiplication).*
7. $1x = x$, *(multiplicative identity).*
8. *For each* $x \neq 0$, *there exists* x^{-1} *such that* $xx^{-1} = 1$.*(existence of multiplicative inverse).*
9. $x(y + z) = xy + xz$.*(distributive law).*

Something which satisfies these axioms is called a field. Linear algebra is all about fields, although in this book, the field of most interest will be the field of complex numbers or the field of real numbers. You have seen in earlier courses that the real numbers also satisfies the above axioms. The field of complex numbers is denoted as $\mathbb{C}$ and the field of real numbers is denoted as $\mathbb{R}$. An important construction regarding complex numbers is the complex conjugate denoted by a horizontal line above the number. It is defined as follows.

$$\overline{a + ib} \equiv a - ib.$$

What it does is reflect a given complex number across the x axis. Algebraically, the following formula is easy to obtain.

$$\begin{aligned}\left(\overline{a + ib}\right)(a + ib) &= (a - ib)(a + ib) \\ &= a^2 + b^2 - i(ab - ab) = a^2 + b^2.\end{aligned}$$

Definition 1.4.2 *Define the absolute value of a complex number as follows.*

$$|a+ib| \equiv \sqrt{a^2+b^2}.$$

Thus, denoting by z the complex number $z = a + ib$,

$$|z| = (z\overline{z})^{1/2}.$$

Also from the definition, if $z = x + iy$ and $w = u + iv$ are two complex numbers, then

$$|zw| = |z|\,|w|.$$

You should verify this. ►

The triangle inequality holds for the absolute value for complex numbers just as it does for the ordinary absolute value.

Proposition 1.4.3 *Let z, w be complex numbers. Then the triangle inequality holds.*

$$|z+w| \leq |z| + |w|, \quad ||z| - |w|| \leq |z-w|.$$

Proof: Let $z = x + iy$ and $w = u + iv$. First note that

$$z\overline{w} = (x+iy)(u-iv) = xu + yv + i(yu - xv)$$

and so $|xu + yv| \leq |z\overline{w}| = |z|\,|w|$.

$$|z+w|^2 = (x+u+i(y+v))(x+u-i(y+v))$$

$$= (x+u)^2 + (y+v)^2 = x^2 + u^2 + 2xu + 2yv + y^2 + v^2$$

$$\leq |z|^2 + |w|^2 + 2|z|\,|w| = (|z| + |w|)^2,$$

so this shows the first version of the triangle inequality. To get the second,

$$z = z - w + w, \; w = w - z + z$$

and so by the first form of the inequality

$$|z| \leq |z-w| + |w|, \; |w| \leq |z-w| + |z|$$

and so both $|z| - |w|$ and $|w| - |z|$ are no larger than $|z - w|$ and this proves the second version because $||z| - |w||$ is one of $|z| - |w|$ or $|w| - |z|$. ■

With this definition, it is important to note the following. Be sure to verify this. It is not too hard but you need to do it.

Remark 1.4.4 *: Let $z = a + ib$ and $w = c + id$. Then $|z - w| = \sqrt{(a-c)^2 + (b-d)^2}$. Thus the distance between the point in the plane determined by the ordered pair (a, b) and the ordered pair (c, d) equals $|z - w|$ where z and w are as just described.*

For example, consider the distance between $(2, 5)$ and $(1, 8)$. From the distance formula this distance equals $\sqrt{(2-1)^2 + (5-8)^2} = \sqrt{10}$. On the other hand, letting $z = 2 + i5$ and $w = 1 + i8$, $z - w = 1 - i3$ and so $(z-w)(\overline{z-w}) = (1-i3)(1+i3) = 10$ so $|z - w| = \sqrt{10}$, the same thing obtained with the distance formula.