Vector Spaces and Linear transformations

Greg Baker

August 4, 2004

Overview

What is a vector? The answer may surprize you. But let's start with the simplest view of a vector. It is an arrow that records distance and direction. By stringing together a sequence of arrows we can provide detailed directions for a journey, or outline an object. It is the way we add arrows to produce a new arrow that really identifies what a vector is. We can incorporate this addition property to other quantities such as velocities, forces, and even functions. What quickly emerges is that it is the linear combination of vectors that allows great diversity in applications and provides deep understanding to the nature of solutions to linear problems. This module starts with the basic description of vectors and then proceeds to elucidate their role in the formation of systems of linear equations.

1 Vector Spaces

It seems intuitive that pointing out a direction and giving a distance translates to an arrow on a map. Further, once an location has been reached, another arrow may be added to continue the journey in a new direction. The final destination is just the direct route obtained by adding the arrows as though they lie on the two sides of a triangle. Figure 1 shows the details. Notice how this is done. First we match off along the vector $\mathbf{x}$. When we reach the end, we restart our journey along $\mathbf{y}$. At that moment we change our point of reference from where we start, the origin, to where we stopped. The end has become the new origin for the vector $\mathbf{y}$. The origin of vectors is often a confusing notion. Just remember they always start from the origin,

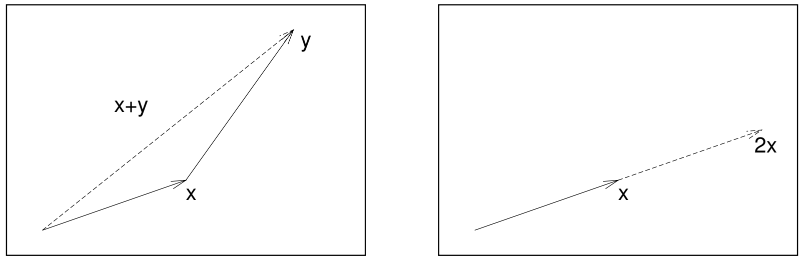

Figure 1: Addition and scalar multiplication of arrows

but they we may change the point of reference at will. Thus the frequent claim that arrows may be displaced. For tangible applications in science and engineering, it is more useful to keep track of our point of reference. So after we reach the end of $\mathbf{x}$, we restart the journey until we reach the end of $\mathbf{y}$. Our journey is the same as if we had simply started at the beginning and traveled along $\mathbf{z} = \mathbf{x} + \mathbf{y}$. It is in this sense that we "add" arrows.

There is another activity we can perform on arrows. We can scale them, that is, change their length. Just imagine giving the direction to travel twice as far as the distance to some location. An example is displayed in Figure 1. Obviously, then, we are free to scale arrows (vectors) at will, and the process is called scalar multiplication.

By combining these two activities, vector addition and scalar multiplication, we can chart courses on maps, mark out boundaries, describe structures, and so on. As an example, let's specify some of the locations of the device shown in Figure 2. The diagram shows a body suspended by wires from two supports. To specify the locations of the points labeled A, B, C, D we must first choose an origin. The choice of the origin is always a matter of convenience, but it is best to choose wisely. For example, we may choose the origin at the point A, in which case the location A is the zero vector – it has no length and its orientation is meaningless. The other points can be represented by arrows from the origin (the point A) to their location. The supports and the body may then be represented by arrows, as shown in Figure 3. The vectors $\mathbf{b}, \mathbf{c}, \mathbf{d}$ give the important locations of the supports, and $\mathbf{e}$ gives the location of the center of the body.

The next step is to understand the balance of forces on the body. There is the force acting straight downwards on the body called the weight $\mathbf{W}$, and

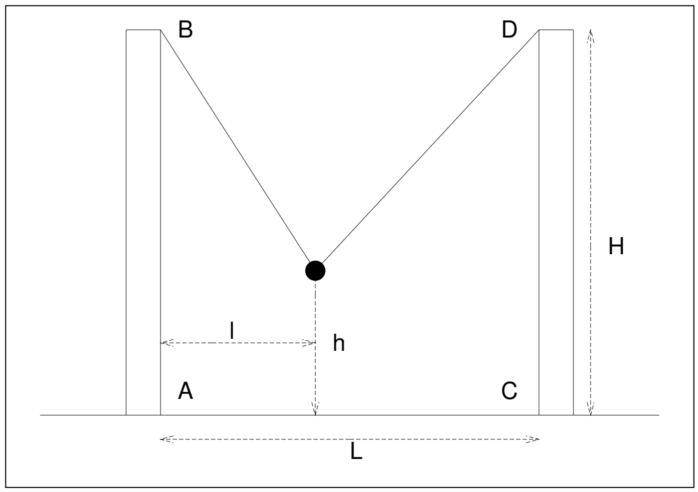

Figure 2: Schematic of a body suspended by two wires

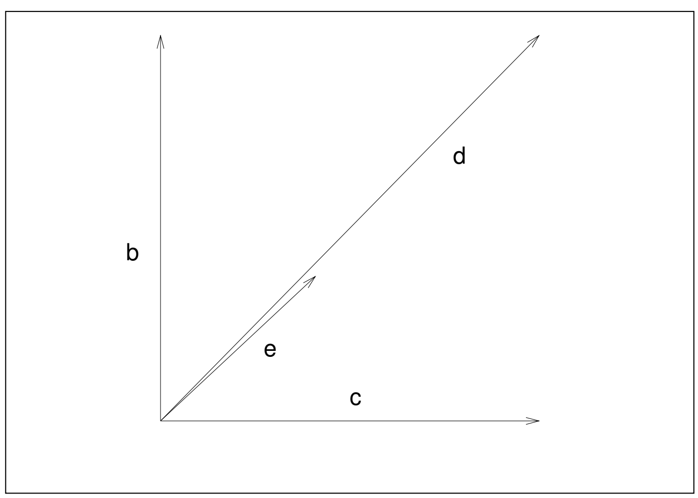

Figure 3: The device represented by arrows

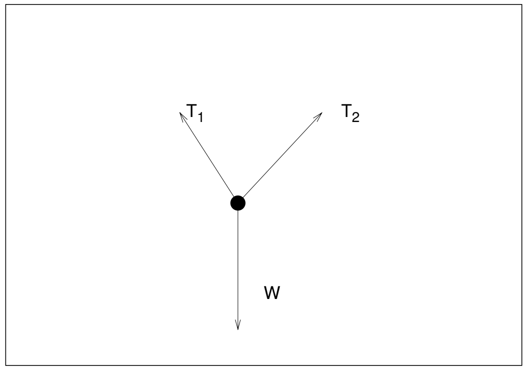

Figure 4: The balance of forces

two forces called tensions that the wires exert on the body to hold it in place. The tensions $\mathbf{T}_1$, $\mathbf{T}_2$ are directed along the wires and their sum must balance the downward pull of the weight. These forces behave like vectors and the force diagram is shown in Figure 4. Notice how the forces are added. First the obvious point of reference is the center of the body. This becomes the origin in the force diagram. The Laws of Physics tell us that the accumulated effect of two forces is a resultant force which may be determined geometrically as follows. Let the two forces be the two sides of a triangle, their lengths and positions reflecting the strength and directions of the forces. The third side gives the strength and direction of the resultant force. This procedure is exactly the same as the process of adding two arrows illustrated in Figure 1. Representing the resultant force as $\mathbf{T}_1 + \mathbf{T}_2$, the static balance of forces requires the resultant force to balance the pull of the weight;

$$\mathbf{T}_1 + \mathbf{T}_2 = \mathbf{W}\,. \tag{1.1}$$

Obviously we may use geometrical considerations to find the tensions once the weight is specified, but such a procedure soon becomes intractible when the structure is more elaborate. Clearly, we need an algebraic method of representation for the vectors, and the representation must be of a nature that can be easily fed into a computer.

The starting point to representing vectors algebraically is to take advantage of vector addition and scalar multiplication. Pick two basis vectors and

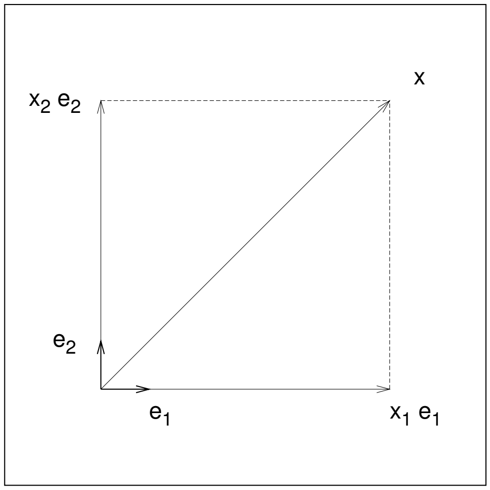

Figure 5: The representation of vectors.

express all other vectors as sums of scalar multiples of these two vectors. The choice of basis vectors is also a matter of convenience, but the obvious choices in our example is a vector $\mathbf{e}_1$ of unit length directed horizontally to the right and a vector $\mathbf{e}_2$ of unit length directed vertically upwards. Then any location vector $\mathbf{x}$ may be expressed as

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 \tag{1.2}$$

The coordinates x_1 and x_2 indicate the multiples of $\mathbf{e}_1$ and $\mathbf{e}_2$ that must be added to produce $\mathbf{x}$. Figure 5 illustrates the representation of $\mathbf{x}$.

As long as everyone knows our choice for $\mathbf{e}_1$ and $\mathbf{e}_2$, all we ever need to tell them to share the location of the vector $\mathbf{x}$ are the coordinates (x_1, x_2) written as an ordered pair of numbers. And, of course, ordered pairs are easily stored in computers. Thus we introduce the short hand notation $\mathbf{x} = (x_1, x_2)$ as long as the basis vectors are clearly understood. Notice $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$, and since $(0, 0)$ indicates the origin, we introduce the zero vector $\mathbf{0} = (0, 0)$ as the vector that locates the origin.

We now have a way to replace the arrows drawn in Figure 3 to represent the supports and weight by a collection of vectors written as ordered pairs:

$$\mathbf{a} = (0,0)\,, \tag{1.3a}$$
$$\mathbf{b} = (0,H)\,, \tag{1.3b}$$
$$\mathbf{c} = (L,0)\,, \tag{1.3c}$$
$$\mathbf{d} = (L,H)\,, \tag{1.3d}$$
$$\mathbf{e} = (l,h)\,. \tag{1.3e}$$

But what have we gained by this representation? Not much unless we know how to add and multiply the vectors in this representation. Fortunately, the consequence of our choice of algebraic representation to the process of vector addition is immediately obvious. Consider two vectors $\mathbf{x}$ and $\mathbf{y}$ with their coordinates (x_1, x_2) and (y_1, y_2) respectively, and draw the sum of the arrows as shown in Figure 6. We add the vector $y_1\mathbf{e}_1$ to the end of vector $x_1\mathbf{e}_1$, then $x_2\mathbf{e}_2$ to the result and finally $y_2\mathbf{e}_2$. We are simply rearranging the order in which we add the arrows. The resulting arrow $\mathbf{x}+\mathbf{y} = (x_1+y_1)\mathbf{e}_1 + (x_2+y_2)\mathbf{e}_2$ shows that we simply add the coordinates to determine the sum of two vectors. In terms of ordered pairs, the addition of vectors is

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)\,. \tag{1.4}$$

We have translated the addition of vectors into the standard arithmetic addition of numbers, something we can instruct a computer to do.

The process of addition illustrated in Figure 6 may be represented algebraically as a series of steps.

$$\begin{aligned}\mathbf{x}+\mathbf{y} &= \big(x_1\mathbf{e}_1 + x_2\mathbf{e}_2\big) + \big(y_1\mathbf{e}_1 + y_2\mathbf{e}_2\big) \\ &= \big(x_1\mathbf{e}_1 + y_1\mathbf{e}_1\big) + \big(x_2\mathbf{e}_2 + y_2\mathbf{e}_2\big) && (1.5\text{a}) \\ &= \big(x_1 + y_1\big)\mathbf{e}_1 + \big(x_2 + y_2\big)\mathbf{e}_2\,. && (1.5\text{b})\end{aligned}$$

In the first step of (1.5a), we have changed the order of the sum to reflect the order in Figure 6. The second step factors the coordinates and the units vectors. These steps may seem very reasonable since they follow the normal steps of algebra when working with numbers. But we are not dealing with numbers, we are dealing with vectors. How do we know that the order of summation of vectors is not important? Because that is the geometric property of the arrows that represent the locations. For example, if we start

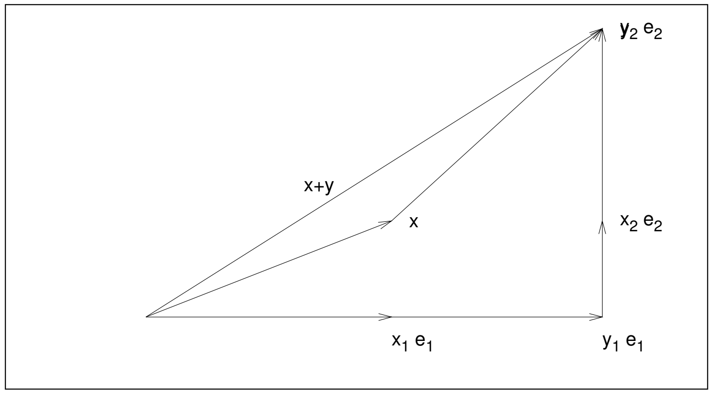

Figure 6: The addition of vectors.

with $\mathbf{x}$ and then add $\mathbf{y}$ we end up at the same location as staring first with $\mathbf{y}$ and then adding $\mathbf{x}$. We may state this result as

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}\,, \tag{1.6}$$

or in terms of the representation of vectors by ordered pairs of coordinates, we have

$$\begin{aligned}(x_1, x_2) + (y_1, y_2) &= (x_1 + y_1, x_2 + y_2) \\ &= (y_1 + x_1, y_2 + x_2)\,,\end{aligned}$$

where we have reverse the order of addition of the coordinates which is allowed because they are just numbers. Alternatively,

$$(y_1, y_2) + (x_1, x_2) = (y_1 + x_1, y_2 + x_2)\,, \tag{1.7}$$

and so the two results agree in accordance with (1.5). The point here is that the algebraic representation of vectors and their sum must conform to the geometric properties of arrows.

Before continuing with an exploration of other algebraic properties of vectors, let us note that scalar multiplication is also easily represented algebraically. Referring to Figure 1, it should be obvious that when we scale an arrow by a factor α say, the result is that the coordinates are both multiplied by the same factor. In other words,

$$\alpha(x_1, x_2) = (\alpha x_1, \alpha x_2) \tag{1.8}$$

Note the obvious. Scalar mutiplication has been transformed into the usual arithmetic multiplication of real numbers. Consequently, we can add and multiply vectors by scalars wth ease, especially with the aid of a computer.

There are further consequences of the representation (1.2) that draw a connection between it and the geometrical properties of an arrow. The length of the arrow may be determined through Pythagoras's theorem. We need a symbol to represent this length, and the mathematician's choices is $\|\mathbf{x}\|$. Another popular choice is r since it is the distance from the origin, but this choice lacks meaning when we wish to measure other vectors such as force. For now, we note that the length of an arrow is

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}\,. \tag{1.9}$$

Notice that the symbol $\|\mathbf{x}\|$ carries the information of which vector is being measured and it is arranged as though it has brackets because we will want to measure the lengths of sums and scalar multiples of vectors. For example, the length of a vector that has been scaled by a factor of α should be just α as long:

$$\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\| \tag{1.10}$$

Obviously, we must take the absolute value of α but otherwise (1.10) has the appearence of moving α outside brackets. Don't be confused, though, the symbol $\|$ is not a bracket – it is just a convenient symbol. More details on the use of $\|$ will be given later.

The angle θ the arrow makes with the horizontal axis (the direction of $\mathbf{e}_1$) is given by

$$\tan\theta = \frac{x_2}{x_1}\,. \tag{1.11}$$

These quantities, the distance and angle, can also be used to express the coordinates as

$$x_1 = \|\mathbf{x}\|\cos(\theta)\,, \qquad x_2 = \|\mathbf{x}\|\sin(\theta) \tag{1.12}$$

So far, we have considered locations in a plane as a matter of convenience. With good imagination or artistic skills we can replace figures, such as Figure 1, with pictures of arrows in three dimensions. Fortunately, we gain no further insight by this generalization, but a certain complication arises in that we must now deal with three coordinates. We can list them as a triplet of ordered numbers which becomes a bit cumbersome. Instead, can list the coordinates as a column. Historically, we prefer columns of numbers if they are long. So we may represent a vector in three dimensions as

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 \tag{1.13a}$$

or

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{1.13b}$$

The addition of column vectors, such as (1.13b), follows the rule of addition for ordered triplets:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} . \tag{1.14a}$$

Scalar multiplication by a real number α is simply

$$\alpha \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{pmatrix} . \tag{1.14b}$$

Well, let us pat ourselves on the back. We have found several ways to represent vectors, as arrows, as ordered lists and as columns of real numbers. But the useful value of these representations is that they apply to other vectors as well, such as velocites and forces, where the length of the arrow now specifies the magnitudes of the vector rather then length. In these cases, we replace the term coordinates with components. For example, we write the velocity $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$ and call v_1 the velocity component in the direction $\mathbf{e}_1$, and similarly for the other components..

To demonstrate the practical value of vector components, let's return to the force balance (1.2) and state it in terms of components. The weight is simply $\mathbf{W} = -W\mathbf{e}_2$, where W gives the magnitude of the weight and the negative sign indicates that the weight acts downwards. To specify the tensions, we need a vector that lies in the direction of the wires. By referring to Figure 3, one of the vectors we need goes from the end of $\mathbf{e}$ to the end of $\mathbf{b}$ while the other goes from the end of $\mathbf{e}$ to the end of $\mathbf{d}$. These vectors are

$$\mathbf{w}_1 = \mathbf{b} - \mathbf{e} = \begin{pmatrix} 0 \\ H \end{pmatrix} - \begin{pmatrix} l \\ h \end{pmatrix} = \begin{pmatrix} l \\ H - h \end{pmatrix} , \tag{1.15a}$$

$$\mathbf{w}_2 = \mathbf{d} - \mathbf{e} = \begin{pmatrix} L \\ H \end{pmatrix} - \begin{pmatrix} l \\ h \end{pmatrix} = \begin{pmatrix} L - l \\ H - h \end{pmatrix} , \tag{1.15b}$$

where we have used column vectors to represent the locations of the wires. Unfortunately, it is not the location of the wires we need but merely a vector of unit length that points along the wires. Well, that's easy enough to create by simply scaling the vectors (1.15) so that the new length is exactly unity. Thus set $\mathbf{t}_1 = \alpha_1 \mathbf{w}_1$ and $\mathbf{t}_2 = \alpha_2 \mathbf{w}_2$ and demand $\|\mathbf{t}_1\| = \|\mathbf{t}_2\| = 1$:

$$\|\mathbf{t}_1\| = \alpha_1 \|\mathbf{w}_1\| = \alpha_1 \sqrt{l^2 + (H-h)^2}\,, \quad \text{so take} \quad \alpha_1 = \frac{1}{\sqrt{l^2 + (H-h)^2}}\,; \tag{1.16a}$$

$$\|\mathbf{t}_2\| = \alpha_2 \|\mathbf{w}_2\| = \alpha_2 \sqrt{(L-l)^2 + (H-h)^2}\,, \quad \text{so take} \quad \alpha_2 = \frac{1}{\sqrt{(L-l)^2 + (H-h)^2}}\,. \tag{1.16b}$$

Now we are in position to state the tensions in the wires as

$$\mathbf{T}_1 = T_1 \mathbf{t}_1 = \frac{T_1}{\sqrt{l^2 + (H-h)^2}} \begin{pmatrix} l \\ H-h \end{pmatrix}, \tag{1.17a}$$

$$\mathbf{T}_2 = T_2 \mathbf{t}_2 = \frac{T_2}{\sqrt{(L-l)^2 + (H-h)^2}} \begin{pmatrix} L-l \\ H-h \end{pmatrix}, \tag{1.17b}$$

where T_1 and T_2 give the strength of the tensions. With these expressions in hand, we may restate (1.1) in more practical terms as

$$\frac{T_1}{\sqrt{l^2 + (H-h)^2}} \begin{pmatrix} x \\ H-h \end{pmatrix} + \frac{T_2}{\sqrt{(L-l)^2 + (H-h)^2}} \begin{pmatrix} L-l \\ H-h \end{pmatrix} = -\begin{pmatrix} 0 \\ W \end{pmatrix}. \tag{1.18}$$

or even as separate equations by balancing components individually,

$$\frac{x}{\sqrt{l^2 + (H-h)^2}} T_1 + \frac{L-l}{\sqrt{(L-l)^2 + (H-h)^2}} T_2 = 0 \tag{1.19a}$$

$$\frac{H-h}{\sqrt{l^2 + (H-h)^2}} T_1 + \frac{H-h}{\sqrt{(L-l)^2 + (H-h)^2}} T_2 = -W \tag{1.19b}$$

One way we might use (1.19) is to specify the location of the weight and determine the needed tension. In this way we could check whether the wires are strong enough. Mathematically, the locations of l and h and the weight W is given along with the design parameters H and L, and we solve the two linear equations (1.19) for T_1 and T_2. The solution to two linear equations is easy. In general, structures are much more complicated and the number of equations and unknowns is quite substantial. So we turn to numerical methods to solve systems of linear equations – see the module on Systems of Linear equations.

Mathematicians don't usually care about the units for the variables found in equations such as (1.19), but they very important in science and engineering. So how do units appear in vectors such as (1.2) or in vector equations as in (1.1)? When we state a vector such as $\mathbf{x}$ we assume it has units appropriate to its physical connection, for example, miles to indicate distance. When we represent vectors as coordinates (or components) as in (1.2), we assign the units to the coefficients. Thus basis vectors, such as $\mathbf{e}_1$, are considered dimensionless and the coefficients, such as x_1, carry the units. This is convenient when we wish to use the same basis vectors to represent velocites so that the component of velocity v_1 has the units appropriate to the chioce of units for x_1. To emphasis further the consistent nature of this viewpoint, look at the nature of the unit vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ that were constructed to represent the tensions in the wires. They are both dimensionless! For example the first component of $\mathbf{t}_1$ is

$$\frac{l}{\sqrt{l^2 + (H - h)^2}} \tag{1.20}$$

which has the ratio of distance over distance.

All the important ideas about vectors have been introduced in this section, and they have been illustrated with familiar examples from physical space and mechanics. There is much more generality in these ideas and we restate them in the next subsection.

1.1 Definitions

The ideas about vectors that have been explored so far can be placed in the context of a mathematical framework. Mathematicians love to make the rules of the game that govern mathematical manipulations and then explore the consequences. This approach has the value that we may identify many more examples of vectors that we can then manipulate according to the rules of the game. So this is what a mathematician considers to be a vector space.

Vector Space:
A vector space is made up of objects that can be added and multiplied by scalars. Not all addition and scalar multiplications are allowed. They must satisfy certain properties. First, the result of the addition must be one of the objects in the vector space. To illustrate this statement consider that the sum of arrows is an arrow and the sum of column vectors is a column vector – see (1.14a). But why this property is important is because we usually add many arrows or column vectors and we can't do that unless the results of each sum is just another object in the set. Similarly, multiplication by a scalar must produce another of the objects in the vector space.

The special properties we demand that addition and scalar multiplication satisfy arise from our desire to perform algebraic steps as in (1.15). For example we must be able to rearrange sums as in (1.15a) and factor out expressions as in (1.15b), but there are more manipulations we wish to perform. Mathematicians have found the smallest list of properties that allow us to perform all the standard algebraic operations, but even so, there are eight of them.

To state all the properties we need, we must choose a notation that represents the objects in the vector space and their addition and scalar multiplication. Let V be the name of the vector space, and let two vectors in it be written simply as $\mathbf{x}$ and $\mathbf{y}$. Of course, in reality we will have a specific representation for these vectors, such as columns of real numbers. Now we must imagine a process of addition. For the moment, let us choose the symbol $\oplus$ to represent the way we add the two vectors $\mathbf{x} \oplus \mathbf{y}$. To demonstrate the application of this notation, consider the usual addition of two column vectors – see (1.14a) – but written with this new symbol:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix} \tag{1.21a}$$

Now it is clear that the plus in the third column is the standard arithmetic plus and the statement tells us how the operation $\oplus$ is to be performed.

We also need a symbol for multiplication by a scalar. Let us pick $\otimes$. We require a scalar to be some real number, α say, but we must still decide how it multiplies a vector $\mathbf{x}$ to produce $\alpha \otimes \mathbf{x}$. The companion to (1.21a) is

$$\alpha \otimes \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \end{pmatrix} \tag{1.21b}$$

Clearly, the definition of the operation $\otimes$ simplifies to standard multiplication of each component of $\mathbf{x}$ by the scalar α.

Beyond the strange choice of symbols, notice what (1.21) really means. We must have a representation for vectors that involve real numbers, in this case, a column of two real numbers. Then we must have a rule to add them (1.21a) and multiply them with a scalar (1.21b) and both rules must use the standard addition and multiplication of real numbers. All this so that at some point we can perform the operations on a computer. Since we also want to use symbolic manipulations, the properties of addition and scalar multiplication should satisfy the following properties:

(**Property** 1) $$\mathbf{x} \oplus \mathbf{y} = \mathbf{y} \oplus \mathbf{x}\,. \tag{1.22a}$$

(**Property** 2) $$(\mathbf{x} \oplus \mathbf{y}) \oplus \mathbf{z} = \mathbf{x} \oplus (\mathbf{y} \oplus \mathbf{z})\,. \tag{1.22b}$$

There is a single vector, called $\mathbf{0}$, for which

(**Property** 3) $$\mathbf{0} \oplus \mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x}\,. \tag{1.22c}$$

There is a vector, called $-\mathbf{x}$, for which

(**Property** 4) $$\mathbf{x} \oplus (-\mathbf{x}) = \mathbf{0}\,. \tag{1.22d}$$

(**Property** 5) $$1 \otimes \mathbf{x} = \mathbf{x} \tag{1.22e}$$

(**Property** 6) $$(\alpha_1 \alpha_2) \otimes \mathbf{x} = \alpha_1 \otimes (\alpha_2 \otimes \mathbf{x}) \tag{1.22f}$$

(**Property** 7) $$\alpha \otimes (\mathbf{x} \oplus \mathbf{y}) = \alpha \otimes \mathbf{x} \oplus \alpha \otimes \mathbf{y} \tag{1.22g}$$

(**Property** 8) $$(\alpha_1 + \alpha_2) \otimes \mathbf{x} = \alpha_1 \otimes \mathbf{x} \oplus \alpha_2 \otimes \mathbf{x} \tag{1.22h}$$

Comments:

1. The first two properties are saying that there is nothing to distinguish vectors in the process of addition. Adding $\mathbf{x}$ to $\mathbf{y}$ is the same as adding $\mathbf{y}$ to $\mathbf{x}$, and it doesn't matter in which order you add a list of vectors; the result is the same. For (1.21a), Property 1 states

$$\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix} = \begin{pmatrix} y_1 + x_1 \\ y_2 + x_2 \end{pmatrix}\,.$$

The results are the same because we can change the order of the standard arithmetic plus. Property 2 states

$$\begin{pmatrix} x_1 + (y_1 + z_1) \\ x_2 + (y_2 + z_2) \end{pmatrix} = \begin{pmatrix} (x_1 + y_1) + z_1 \\ (x_2 + y_2) + z_2 \end{pmatrix}$$

and the results are the same because of the properties of the arithmetic plus.

2. Property 3 is really demanding that the "origin," as represented by **0**, must be in the set of vectors. We must have a starting point in defining vectors. For (1.21a), the origin is

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which should be no surprise. Obviously,

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ,$$

and there is only one vector **0** which has this property.

3. In Property 4, $-\mathbf{x}$ is a compound symbol represent just one vector. It is the vector which retraces our steps so that we return to the origin. Obviously, then, it is a vector that has the same length as $\mathbf{x}$, but points in the opposite direction. How do we find the components of $-\mathbf{x}$? Let's illustrate how for (1.21a). Let's call y_1 and y_2 the components of $-\mathbf{x}$. We must have

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Our choice is now clear: $y_1 = -x_1$ and $y_2 = -x_2$. In the same way that we interpret the addition of a negative number as subtraction, we may write

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ominus \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

In other words, we may use

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ominus \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \end{pmatrix} . \tag{1.23}$$

4. Property 5 says that if we scale a vector by unity then the vector should't change. This is obviously true for (1.21b):

$$1 \otimes \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

5. On the other hand, if we double a vector then triple it, we want the result to be the same as scaling the vector by a factor of six – hence Property 6. For (1.21b),

$$(2 \times 3) \otimes \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6x_1 \\ 6x_2 \end{pmatrix} .$$

Properties 5 and 6 ensure that scaling vectors makes good sense in that proportionality will be maintained.

6. Properties 7 and 8 couple together the operations of vector addition and scalar multiplication. They tell us what to do when combinations of the operations arise. For (1.12), Property 7 states

$$\begin{pmatrix} \alpha(x_1 + y_1) \\ \alpha(x_2 + y_2) \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \end{pmatrix} \oplus \begin{pmatrix} \alpha y_1 \\ \alpha y_2 \end{pmatrix} ,$$

while Property 8 states

$$\begin{pmatrix} (\alpha_1 + \alpha_2)x_1 \\ (\alpha_1 + \alpha_2)x_2 \end{pmatrix} = \begin{pmatrix} \alpha_1 x_1 \\ \alpha_1 x_1 \end{pmatrix} \oplus \begin{pmatrix} \alpha_2 x_1 \\ \alpha_2 x_2 \end{pmatrix} .$$

7. With these basic properties, we may multiply and add vectors to our heart's content, no matter what order the operations and in all cases the result will be a vector.

All this formalism may seem unnecessary. After all, we all know that vectors in a plane just need two components to specify their location, and we need three components if we must locate a vector in a volume. So, let's just use them and get on with it. However, special circumstances may arise which may cause difficulties. Let us consider a special example that illustrates how a vector space may arise in three-dimensions where the addition and scalar multiplication of vectors may be unexpected.

Suppose we are interested in physical space where the origin is placed on the first floor of a building. We wish to place furniture on the second floor of the building by specifying the location *relative to a fixed marker on the second floor*. For example, place a table three feet from the entrance to the dining room. To make the mathematics as simple as possible, let our marker be placed at

$$\mathbf{s} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} . \tag{1.24}$$

In other words, the marker on the second floor is placed one unit (about ten feet) above the origin on the first floor. Now all locations on the second floor can be specified as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} , \tag{1.25}$$

once we have chosen the directions corresponding to the components x_1 and x_2 to be the same as the coordinate system we have chosen for the three-dimensiopnal space. Mathematically, we consider all vectors defined by (1.25) to lie in a plane.

Now we wish to add two vectors of the form (1.25) so that the new location is still on the second floor. It is tempting to use the usual addition (1.14a) of vectors in three-dimensions. If we do, the result will be

$$\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ 2 \end{pmatrix} ,$$

which is a vector pointing to a location on the third floor! That's no good. It is the addition of the third components that shift the result to the third floor. With that insight, we try again.

Let V_2 be the set of all vectors of the form (1.25). The addition of two vectors $\mathbf{x}$ and $\mathbf{y}$ from V_2 will be given by

$$\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \oplus \begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ 1 \end{pmatrix} , \tag{1.26a}$$

and scalar multiplication by

$$\alpha \otimes \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ 1 \end{pmatrix} . \tag{1.26b}$$

Notice in both cases that the result is a vector belonging to V_2. Although tedious, it is easy to check that these definitions of addition and scalar multiplication satisfy all the Properties (1)-(8). Along the way, we will discover that $\mathbf{0} = \mathbf{s}$ and that

$$-\mathbf{x} = \begin{pmatrix} -x_1 \\ -x_2 \\ 1 \end{pmatrix} .$$

There is another reason it is useful to distill the ideas of vectors into a more abstract formulation. We may find many other surprising examples of vectors that share the same properties. A simple example is the spaces of polynomials. To be specific, let's define P_n as the set of all polynomials whose highest power is $n-1$ (called the degree of the polynomial). The reason for the shift in n is that there are n coefficients that describe a polynomial of degree $n-1$. For example, P_4 is the set of all cubic polynomials. A general way to express a cubic polynomial is

$$p_4(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \,, \tag{1.27}$$

and there are four coefficients (a_0, a_1, a_2, a_3). It is standard to associate the subscript on the coefficients with the power of x because we can then write a general vector from P_n in the compact form

$$p_n(x) = \sum_{k=0}^{n} a_k x^k \,.$$

We now claim P_n is a vector space under the standard rules for adding polynomials and multiplying them with real numbers. Specifically, consider P_4. Pick two "vectors" from P_4,

$$\mathbf{x} = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \,, \tag{1.28a}$$
$$\mathbf{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3 \,. \tag{1.28b}$$

Their addition is given as

$$\mathbf{x} \oplus \mathbf{y} = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + (a_3 + b_3)x^3 \,. \tag{1.29a}$$

We simply add the coefficients of the like powers of x. The scalar multiplication of a polynomial is

$$\alpha \otimes \mathbf{x} = (\alpha a_0) + (\alpha a_1)x + (\alpha a_2)x^2 + (\alpha a_3)x^3 \,. \tag{1.29b}$$

Each coefficient is multiplied by the scalar. The result has been written with extra parentheses to emphasize that the new coefficients are just multiplied by the scalar α. The astute observer will notice that the addition and scalar multiplication look exactly like the addition and scalar multiplication of column vectors with four components once the coefficents are written as a column. Obviously,

$$\mathbf{0} = 0 + 0x + 0x^2 + 0x^3 \,, \tag{1.30a}$$

$$-\mathbf{x} = -a_0 - a_1 x - a_2 x^2 - a^3 x^3 \,, \tag{1.30b}$$

and all the Properties (1) - (8) are satisfied.

There are two similarities in the vector spaces generated by the addition and scalar multiplication of arrows and of polynomials. From (1.2), we can represent any arrow as a linear combination of two basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Subsequently, we simply place the components x_1 and x_2 into a column, which we call a column vector. By contemplating the nature of the polynomial (1.27), we notice we may express it as a linear combination of four "basis vectors"

$$p_4(x) = a_0\mathbf{e}_0 + a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3 \,. \tag{1.31}$$

where

$$\mathbf{e}_0 = 1\,, \qquad \mathbf{e}_1 = x\,, \qquad \mathbf{e}_2 = x^2\,, \qquad \mathbf{e}_3 = x^3\,. \tag{1.32}$$

Each basis vector is a simple polynomial from P_4. Obviously we may represent any polynomial $p \in P_4$ (this is the mathematical shorthand that says the polynomial belong to the set P_4) as a column vector with the coefficients ordered from top to bottom,

$$p \equiv \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \,. \tag{1.33}$$

The idea of expressing a vector from a vector space as a linear combination of basis vectors is extremely useful since it immediately results in the ability to express the coefficients (or components) as column vectors, and column vectors can be stored and manipulated on computers. We will explore this idea further in the next section.

2 Linear Combinations.

The truth of the matter is that the identification and creation of vector spaces is often much simpler than the definitions in section 1.1 would imply. And the reason is that linear combinations are the fundamental tools in vector spaces. When we see a linear combination, we should expect an underlying vector space.

To illustrate my point, suppose we are a small grocery that sells bananas, apples, oranges and peaches. A typical purchase may be recorded as three bananas and two apples and one orange and four peaches. Let's begin to restate this purchase order with the shorthand of mathematics. We'll introduce $\mathbf{p}$ to represent the purchase order and write it as

$$\mathbf{p} = 3 \times \text{bananas} + 2 \times \text{apples} + \text{orange} + 4 \times \text{peaches}\,. \tag{2.1}$$

Notice the similarity of this statement with that of a polynomial (1.27). Just as we can't add x with x^2 in general, we can't add apples and oranges, but we can add two purchase orders together and we can multiple a purchase order with a scale factor. Obviously, the addition of two purchase orders

$$\mathbf{p} = p_1 \times \text{bananas} + p_2 \times \text{apples} + p_3 \times \text{oranges} + p_4 \times \text{peaches}\,, \tag{2.2a}$$
$$\mathbf{q} = q_1 \times \text{bananas} + q_2 \times \text{apples} + q_3 \times \text{oranges} + q_4 \times \text{peaches} \tag{2.2b}$$

should be

$$\begin{aligned}\mathbf{p} \oplus \mathbf{q} = {} & (p_1 + q_1) \times \text{bananas} + (p_2 + q_2) \times \text{apples} \\ & + (p_3 + q_3) \times \text{oranges} + (p_4 + q_4) \times \text{peaches}\,,\end{aligned} \tag{2.2c}$$

and the multiplication by a scalar should be

$$\begin{aligned}\alpha \otimes \mathbf{p} = {} & (\alpha p_1) \times \text{bananas} + (\alpha p_2) \times \text{apples} \\ & + (\alpha p_3) \times \text{oranges} + (\alpha p_4) \times \text{peaches}\,.\end{aligned} \tag{2.3}$$

Clearly, it is tedious to keep refering to the quantities "bananas,", "apples," etc., especially when this list would be very long in practice. The obvious shorthand is to introduce an ordered list,

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}\,. \tag{2.4}$$

In general, lists such as a purchase order are very common. More examples include students grades, parts in an inventory, seat reservations, and on and on. By simply choosing a specific order, we may list quantities as a column

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{2.5}$$

with the number of entries n matching the number of items in the list. Then it is natural to add two lists or multiply lists as

$$\mathbf{x} \oplus \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} , \qquad \alpha \otimes \mathbf{x} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix} . \tag{2.6}$$

Since this situation is so common, we call (2.5) a column vector belong to the vector space R^n.

Just as we introduce unit vectors for arrows as in (1.2), we may introduce unit vectors for lists. For example, the unit vector that identifies "bananas" in our grocery list is

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} . \tag{2.7a}$$

The order for one banana would read $\mathbf{p} = \mathbf{e}_1$. Similarly, we may introduce unit vectors for "apples," "oranges" and "peaches" as

$$\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} , \qquad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} , \qquad \mathbf{e}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} . \tag{2.7b}$$

Then we may write our grocery list as

$$\mathbf{p} = p_1\mathbf{e}_1 + p_2\mathbf{e}_2 + p_3\mathbf{e}_3 + p_4\mathbf{e}_4 \tag{2.7c}$$

This dual representation of column vectors, (2.4) and (2.7c), proves very convenient as we shall see.

Competition from big supermarkets is driving our fruit business into the ground. So we decide to sell fruit baskets as special orders on the internet. We plan to offer three sizes: a small basket containing two bananas and an apple; a medium basket containing four bananas, two apples an orange; and a large basket containing six bananas, two apples, two oranges and a peach. Using our previous ordered list, we may represent these three baskets as column vectors

$$\mathbf{a}_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{a}_2 = \begin{pmatrix} 4 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{a}_3 = \begin{pmatrix} 6 \\ 2 \\ 2 \\ 1 \end{pmatrix}. \tag{2.8}$$

An order from the local hospital might look like ten small baskets, four medium baskets and 2 large baskets. Mathematically, we may represent the total order as

$$\mathbf{p} = 10\mathbf{a}_1 + 4\mathbf{a}_2 + 2\mathbf{a}_3 \tag{2.9a}$$

$$= 10 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + 4 \begin{pmatrix} 4 \\ 2 \\ 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 6 \\ 2 \\ 2 \\ 1 \end{pmatrix} \tag{2.9b}$$

$$= \begin{pmatrix} 10 \times 2 + 4 \times 4 + 2 \times 6 \\ 10 \times 1 + 4 \times 2 + 2 \times 2 \\ 10 \times 0 + 4 \times 1 + 2 \times 2 \\ 10 \times 0 + 4 \times 0 + 2 \times 1 \end{pmatrix} \tag{2.9c}$$

$$= \begin{pmatrix} 48 \\ 22 \\ 8 \\ 2 \end{pmatrix}. \tag{2.9d}$$

2.1 Sub-spaces

Each line in (2.9) is very important. The first line (2.9a) illustrates a linear combination, which, in general, is simply the addition of multiples of vectors. Linear combinations have appeared several times already in these notes, for example, in (1.2), (1.31) and (2.7c). Linear combinations and vector spaces are closely related. The rules that govern vector spaces are just the rules needed to form linear combinations and manipulate them. On the other hand, if we consider all possible linear combinations we create a vector space. Some examples will illustrate how this can be done.

The first example is the use of the unit arrows $\mathbf{e}_1$ and $\mathbf{e}_2$ in (1.2). By considering every possible linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$ we obtain all vectors in the plane. There is a mathematical shorthand for considering all possible linear combinations of a set of vectors. We call it a span. Specifically, suppose $\mathbf{v}_j$, $j = 1, \ldots n$ is a list of vectors from some vector space and α_j are real numbers. Then we write a span as a set of vectors

$$\begin{aligned}\mathrm{span}\big\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\big\} &\equiv \big\{\mathbf{x} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n \,|\, \text{for all } \alpha_j \in R\big\} \\ &= \big\{\mathbf{x} = \sum_{j=1}^{n} \alpha_j \mathbf{v}_j \,|\, \text{for all } \alpha_j \in R\big\}\,. \end{aligned} \tag{2.10}$$

Another example of a span is the way we may express all cubic polynomials as

$$\mathrm{span}\big\{1, x, x^2, x^3\big\}\,, \tag{2.11}$$

which is the same as the vector space P_4 – see (1.31).

A better appreciation of the value of a span emerges from our example of fruit baskets. First let us consider the vector space $\mathcal{P}$ generated by all possible purchase orders as given by (2.7c). However, for the purchase of fruit baskets, not all linear combinations are allowed. Instead we must insist that only linear combinations of the form (2.9a) are allowed. The vector space $\mathcal{F}$ of purchase orders for fruit baskets is given by

$$\mathcal{F} = \mathrm{span}\big\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\big\}\,, \tag{2.12}$$

where $\mathbf{a}_j$ are the vectors representing fruit baskets defined in (2.8).

How do we know $\mathcal{F}$ is a vector space? Well, first of all we know how to add purchase orders and we know how to multiply them with a scalar. Ah, but we must be able to multiply them by any scalar, including negative numbers. Obviously, we will interpret a negative purchase order as a returned order. Next we must confirm that all the properties (1.22) apply. We can tediously go through all the properties and check them one by one. But there is a simple way to see that they must hold. The vectors $\mathbf{a}_j$ all belong to the column vector space R^4 and thus any addition or multiplication by a scalar must be true for them. Consequently, any linear combination of $\mathbf{a}_j$ is just the addition and scalar multiplication of column vectors, which in turn satisfy the properties (1.22).

But there remains two further important tests:

1. The result of the addition of two vectors must result in a vector belonging to $\mathcal{F}$.

2. The multiplication of a vector by a scalar must result in a vector belonging to $\mathcal{F}$.

We call these two tests the *closure properties*.

By the very nature of spans, we find that they automatically satisfy the closure properties. Here's why. Take any two vectors from $\mathcal{F}$;

$$\mathbf{p}_1 = \alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \alpha_3\mathbf{a}_3\,, \tag{2.13a}$$

$$\mathbf{p}_2 = \beta_1\mathbf{a}_1 + \beta_2\mathbf{a}_2 + \beta_3\mathbf{a}_3\,. \tag{2.13b}$$

When we say any vectors from $\mathcal{F}$ we mean we must allow all possible choices for α_j and β_j. Obviously,

$$\mathbf{p}_1 + \mathbf{p}_2 = (\alpha_1 + \beta_1)\mathbf{a}_1 + (\alpha_2 + \beta_2)\mathbf{a}_2 + (\alpha_3 + \beta_3)\mathbf{a}_3 \tag{2.14}$$

is any vector belonging to $\mathcal{F}$ since it is a linear combination of $\mathbf{a}_j$. The first closure property is satisfied. Now multiply $\mathbf{p}_1$ by the scalar ρ. The result

$$\rho\mathbf{p}_1 = (\rho\alpha_1)\mathbf{a}_1 + (\rho\alpha_2)\mathbf{a}_2 + (\rho\alpha_3)\mathbf{a}_3 \tag{2.15}$$

also belongs to $\mathcal{F}$ and so the second closure property is satisfied.

Let's capture what we have done with some simple mathematical language. The vector space $\mathcal{F}$ is an example of a more general definition of a subset of a vector space V,

$$\mathcal{S} \equiv \big\{\mathbf{v} \in V \mid \text{some restriction}\big\}\,. \tag{2.16}$$

If $\mathcal{S}$ satisfies the closure properties then it is a vector space and we call it a *sub-space* of V. For our example, the vector space is R^4 and the restriction is $\mathbf{v} \in \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. In other words, we can write the vector space of fruit baskets as

$$\mathcal{F} \equiv \Big\{\mathbf{p} \in R^4 \mid \mathbf{p} \in \text{span}\big\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\big\}\Big\}\,. \tag{2.17}$$

Not all restrictions ensure that $\mathcal{S}$ is a sub-space, or even a vector space at all. The first example is the subset

$$\mathcal{S} \equiv \left\{ \mathbf{x} \in R^2 \,|\, x_1 \geq 0 \right\} . \tag{2.18}$$

It is neither a sub-space nor a vector space because the closure properties are not satisfied. For example, let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in R^2$$

with $x_1 \geq 0$ and multiply it with -1. The result is

$$\begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix}$$

and the first component is obviously negative. So the result does not belong to $\mathcal{S}$. The geometrical significance of $\mathcal{S}$ is that the restriction picks out only the right half region of the plane and this region is not a vector space.

This next example demonstrates what is a good subspace for the plane R^2.

$$\mathcal{S} \equiv \left\{ \mathbf{x} \in R^2 \,|\, x_2 = 2x_1 \right\} . \tag{2.19}$$

First note the restriction is specifying a straight line passing through the origin. Lest you be confused, let's state this restriction in more familiar notation: let $y = x_2$ and $x = x_1$, then the restriction states $y = 2x$. For $\mathcal{S}$ to be a sub-space, it must satisfy the closure properties. Let's pick two vectors that belong to $\mathcal{S}$. We may express them as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ 2x_1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y_1 \\ 2y_1 \end{pmatrix} . \tag{2.20a}$$

Their addition results in

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ 2x_1 + 2y_1 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ 2(x_1 + y_1) \end{pmatrix} . \tag{2.20b}$$

Clearly the second component is just twice the first and the result belongs to $\mathcal{S}$. By multiplying $\mathbf{x}$ with the scalar α we find

$$\alpha \mathbf{x} = \begin{pmatrix} \alpha x_1 \\ 2\alpha x_1 \end{pmatrix} \tag{2.20c}$$

and the result clearly belongs to $\mathcal{S}$. Incidentally, if we try the restriction $x_2 = 2x_1 + 1$, the closure properties are not satisfied. In other words, straight lines are subspaces if and only if they pass through the origin.

What do the subspaces for R^3 or even R^n look like? The answer is simple but it is best to revisit (2.19) first to express the restriction in another way. Let's write the vector $\mathbf{x}$ in (2.20a) as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ 2x_1 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} . \tag{2.21a}$$

If we call

$$\mathbf{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} , \tag{2.21b}$$

and rename $x_1 = \alpha$ (it is just a scalar!), then $\mathbf{x} = \alpha\mathbf{a}$. In other words, $\mathbf{x} = \text{span}\{\mathbf{a}\}$. We simply pick a vector that lies in the direction of the straight line, then all possible multiplies of it will mark out the straight line. How simple! How elegant!

So how do we specify a straight line passing through the origin in R^3? Simply by finding a vector $\mathbf{a}$ in the direction of the line and considering the span of $\mathbf{a}$. To specify a plane passing through the origin of R^3, we must find two vectors lying in the plane (they must not point in the same direction). Then the span of these two vectors marks out a plane. We illustrate this point in Figure 7. Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be our choice of vectors in the plane. The any vector in the plane can be reached by an appropriate combination of multiples of $\mathbf{a}_1$ and $\mathbf{a}_2$. The subspace is defined by

$$\mathcal{S} \equiv \left\{ \mathbf{v} \in R^3 \,|\, \mathbf{v} \in \text{span}\{\mathbf{a}_a, \mathbf{a}_2\} \right\} . \tag{2.22}$$

Not all subsets that are vector spaces are subspaces. Recall the vector space we constructed to represent vectors on the second floor of a building (1.25). We found that the subset

$$\mathcal{S} \equiv \left\{ \mathbf{x} \in R^3 \,|\, x_3 = 1 \right\} \tag{2.23}$$

is a vector space provided we define addition by (1.26a) and multiplication by a scalar by (1.26b). But $\mathcal{S}$ is not a subspace of R^3; the closure properties are not satisfied because we would have to use the standard definitions for addition and multiplication by a scalar and not the special ones as given in (1.26).

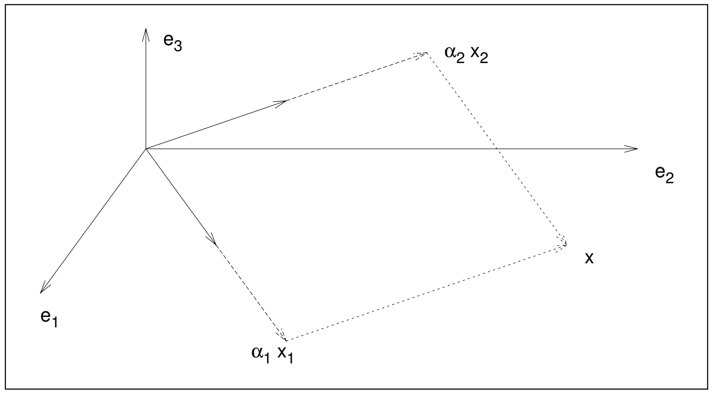

Figure 7: A plane as a subspace of R^3.

In summary, linear combinations allows us to define and represent subspaces. This is very useful when we have lists of items that contain combinations, such as our fruit baskets.

2.2 Matrix Multiplication

Let's us now turn attention to (2.9c), the step that shows how to calculate the total number of fruit items given the number of fruit baskets. First, we can imagine recording the number for each type of fruit basket as a column vector. Specifically, we could simply record

$$\mathbf{p}_a = \begin{pmatrix} 10 \\ 4 \\ 2 \end{pmatrix} \tag{2.24a}$$

for the purchase order (2.9a). We have called this column vector $\mathbf{p}_a$ to remind ourselves that it represents the purchase order in terms of the types of fruit baskets. We can also express the purchase order in terms of the items of fruit (2.9d)

$$\mathbf{p}_e = \begin{pmatrix} 48 \\ 22 \\ 8 \\ 2 \end{pmatrix} . \tag{2.24b}$$

Here we use the notation $\mathbf{p}_e$ to remind ourselves this is the purchase order in terms of the standard unit vectors, the items of fruit. We have dual represents for the same purchase order depending on whether we are thinking in terms of fruit items or fruit baskets. The step that connects the two representations is (2.9c).

The crucial information that is needed for step (2.9c) is the representation of a fruit basket in terms of the items of fruit. This information tells us how to relate fruit baskets to fruit items. So we list all the column vectors in the correct order and form a table

$$A_{a\to e} \equiv \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3\end{bmatrix} \tag{2.24c}$$

$$\equiv \begin{bmatrix}2 & 4 & 6\\ 1 & 2 & 2\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}, \tag{2.24d}$$

and we call the table a matrix. The subscript $a \to e$ reminds us that we are converting information about the number of fruit baskets to the number of fruit items. We can capture the essence of step (2.9c) by a diagram

$$\mathbf{p}_a \quad \xrightarrow[A_{a\to e}]{} \quad \mathbf{p}_e\,. \tag{2.25a}$$

Algebraically, we would like to express the step as

$$\mathbf{p}_e = A_{a\to e}\,\mathbf{p}_a \tag{2.25b}$$

because it implies that the matrix $A_{a\to e}$ "acts" on $\mathbf{p}_a$ to produce $\mathbf{p}_e$. Since $A_{a\to e}\,\mathbf{p}_a$ must be equivalent to step (2.9c), we must have

$$\begin{bmatrix}2 & 4 & 6\\ 1 & 2 & 2\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}\begin{pmatrix}10\\ 4\\ 2\end{pmatrix} \equiv \begin{pmatrix}10\times 2 + 4\times 4 + 2\times 6\\ 10\times 1 + 4\times 2 + 2\times 2\\ 10\times 0 + 4\times 1 + 2\times 2\\ 10\times 0 + 4\times 0 + 2\times 1\end{pmatrix}. \tag{2.25c}$$

On the left-hand side, it appears as though we are multiplying a matrix and a vector; hence it is called matrix-vector multiplication.

The rules governing matrix-vector multiplication are revealed by the pattern in the numbers. To obtain the topmost entry in the result we take the first row of the matrix and multiply each column with the associated entry in the column of $\mathbf{p}_a$ and add the results. We repeat this process for each row of the matrix.

In general, suppose x_j are the components of a column vector $\mathbf{x} \in R^n$ and a_{ij} are the entries in a matrix $A \in R^{m\times n}$. The symbol $R^{m\times n}$ stands for the set of all matrices with entries that are real numbers and identifies the size of the matrix: there are m rows and n columns. The subscripts i and j refer to the row number and the column number of the entry a_{ij}. If we consider the j column of A as a vector $\mathbf{a}_j$, then its ith component $(\mathbf{a}_j)_i = a_{ij}$. It seems that the order of the indies are reversed, but it is all just a matter of style. What is important is the rule for matrix-vector multiplication. If $\mathbf{y} = A\mathbf{x}$ has components y_i, then

$$y_i = \sum_{j=1}^{n} a_{ij}x_j \quad \text{for } i = 1, \dots m\,. \tag{2.26a}$$

Note that i ranges of over the number of rows of A and defines the number of rows (components) of the column vector $\mathbf{y}$. The sum extends over the number of columns of A and the number of rows (components) of the column vector $\mathbf{x}$. Matrix-vector multiplication is not defined if the number of columns of A do not match the number of rows of $\mathbf{x}$.

If we regard the columns of A as vectors $\mathbf{a}_j$, then (2.26a) is equivalent to

$$\mathbf{y} = \sum_{j=1}^{n} x_j\mathbf{a}_j\,, \tag{2.26b}$$

which is just a linear combination of the columns vectors of A. Often, we proceed in the opposite way; we interpret a linear combination as a matrix-vector multiplication.

There are certain important consequences of the definition of matrix-vector multiplication. Suppose we wish to multiply a matrix with a scalar multiple of a vector, specifically $\mathbf{z} = A(\alpha\mathbf{x})$. Then

$$\begin{aligned} z_i &= \sum_{j=1}^{n} a_{ij}(\alpha x_j) \\ &= \alpha \sum_{j=1}^{n} a_{ij}x_j\,, \end{aligned} \tag{2.27}$$

or $\mathbf{z} = \alpha A\mathbf{x}$. Due to the nature of summation, there are several other properties of matrix-vector multiplication. We list two important ones:

$$A(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha A\mathbf{x} + \beta A\mathbf{y}\,, \tag{2.28a}$$

$$(\alpha A + \beta B)\,\mathbf{x} = \alpha A\mathbf{x} + \beta B\mathbf{x}\,. \tag{2.28b}$$

Here, α and β are scalars, $A, B \in R^{m\times n}$ and $\mathbf{x}, \mathbf{y} \in R^n$. In particular, the first property (2.28a) indicates how we would process multiple purchase orders to determine the total fruit items required.

In summary, we see that linear combinations may be stated as a matrix-vector multiplication. This view encourages the interpretation that a matrix acts on or transforms a vector into another. The example of fruit baskets and fruit items demonstrates clearly the value of such an interpretation. All that we need to do is form a matrix with the column vectors that convert each fruit basket into a list of fruit items, and then form a column vector of the number of fruit baskets in the purchase order. The subsequent multiplication of the matrix with this vector produces a new vector with the total number of each fruit item. Our action has taken one linear combination into a new linear combination, and we call such an action a *linear transformation*. In the next section, we explore several more important vector spaces and linear transformation between them.

3 Linear Transformations

As the example of fruit baskets and fruit items in the previous section demonstrates, linear transformation describe how a vector written as a linear combination is "transformed" into another vector also written as a linear combination. A classical example of a linear transformation is the method of relating the components of a vector in three-dimensional space in one frame of reference to another. For a specific example, suppose the device illustrated in Figure 2 is to be placed on a slope, How then should we express the force balance in (1.1) in this new configuration?

Clearly, the first step is to find a way to relate the two frames of reference. Suppose we inroduce new unit vectors $\mathbf{a}_1$, $\mathbf{a}_2$ to establish the new frame of reference that identiifies the device on the slope as illustrated in Figure 8. It is easy to express the unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ as linear combination of $\mathbf{e}_1$ and

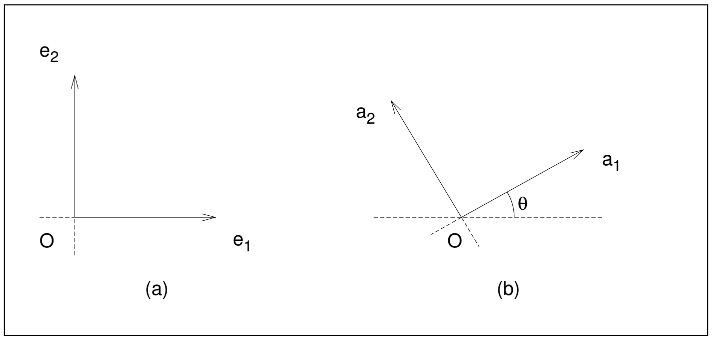

Figure 8: The orientation of the unit vectors for the the device on: a) a flat surface; b) a slope of angle θ.

$\mathbf{e}_2$:

$$\mathbf{a}_1 = \cos(\theta)\,\mathbf{e}_1 + \sin(\theta)\,\mathbf{e}_2\,, \tag{3.1a}$$

$$\mathbf{a}_2 = -\sin(\theta)\,\mathbf{e}_1 + \cos(\theta)\,\mathbf{e}_2\,. \tag{3.1b}$$

Or, equivalently, we may express them in terms of column vectors

$$\mathbf{a}_1 = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}\,, \quad \mathbf{a}_2 = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}\,. \tag{3.2}$$

We can express a vector in the new frame as

$$\mathbf{x}_a = \alpha_1\,\mathbf{a}_1 + \alpha_2\,\mathbf{a}_2 \tag{3.3a}$$

$$= \alpha_1 \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} + \alpha_2 \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} \tag{3.3b}$$

$$= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \tag{3.3c}$$

$$= \begin{pmatrix} \cos(\theta)\,\alpha_1 - \sin(\theta)\,\alpha_2 \\ \sin(\theta)\,\alpha_1 + \cos(\theta)\,\alpha_2 \end{pmatrix}\,. \tag{3.3d}$$

Now we can express the same vector in terms of the old frame as

$$\mathbf{x}_e = x_1\,\mathbf{e}_1 + x_2\,\mathbf{e}_2\,, \tag{3.4a}$$

or as

$$\mathbf{x}_e = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} . \tag{3.4b}$$

By comparing (3.3d) with (3.4b), we have

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta)\,\alpha_1 - \sin(\theta)\,\alpha_2 \\ \sin(\theta)\,\alpha_1 + \cos(\theta)\,\alpha_2 \end{pmatrix} \tag{3.4c}$$

$$= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} . \tag{3.4d}$$

Before emphasizing the importance and usefulness of these results, let us note the direct analogy with (2.9) and (2.25). A linear combination (3.3a) has been expressed as a matrix-vector multiplication (3.3c). The coordinates (components) of $\mathbf{x}_e$ are obtained by a matrix-vector multiplication (3.4d) on $\mathbf{x}_a$. So to complete the analogy, define the transformation matrix

$$\mathcal{T}_{a\to e} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \tag{3.5a}$$

$$= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} , \tag{3.5b}$$

then (3.4d) is the same as

$$\mathbf{x}_e = \mathcal{T}_{a\to e}\,\mathbf{x}_a \tag{3.5c}$$

To demonstrate the practical importance of (3.5c), let's locate the position of the weight. Since $\mathbf{a}_1$ and $\mathbf{a}_2$ are the unit vectors fixed with the device, we simply replace the unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ used to specify $\mathbf{e}$ – see (1.3e) – by $\mathbf{a}_1$ and $\mathbf{a}_2$ respectively. Thus,

$$\mathbf{e}_a = l\mathbf{a}_1 + h\mathbf{a}_2 . \tag{3.6a}$$

where we introduce the subscript a to remind us that the coordinates (l, h) are in terms of the new frame. If we wish to know the location of the weight in standard horizontal ($\mathbf{e}_1$) and vertical ($\mathbf{e}_2$) coordinates then we must apply the transformation matrix $\mathcal{T}_{a\to e}$ to obtain

$$\mathbf{e}_e = \big[l\cos(\theta) - h\sin(\theta)\big]\,\mathbf{e}_1 + \big[l\sin(\theta) + h\cos(\theta)\big]\,\mathbf{e}_2 \tag{3.6b}$$

All of the position vectors (1.3) are specified in the same way, which means the expressions for the tensions in the wires are exactly the same as (1.17), except, of course, the the components are referred to $\mathbf{a}_1$ and $\mathbf{a}_2$. But there is a difference in the balance of forces (1.1) because the weight $\mathbf{W} = -W\mathbf{e}_2$ must point downwards and we must not replace $\mathbf{e}_2$ by $\mathbf{a}_2$ in this expression. Geometrical considerations based on Figure 8 lead to

$$\mathbf{e}_2 = \cos(\theta)\,\mathbf{a}_1 + \sin(\theta)\,\mathbf{a}_2\,. \tag{3.7}$$

By the way, the same result can be obtained by considering the system of equations (3.4c) and setting $x_1 = 0$ and $x_2 = 1$ which are the coordinates of $\mathbf{e}_2$. The solution for α_1 and α_2 give the coordinates of $\mathbf{e}_2$ in the new frame of reference. Thus the weight

$$\mathbf{W} = -W\begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix} \tag{3.8}$$

in the new coordinate system. The force balance in the new frame of reference is

$$\frac{T_1}{\sqrt{l^2+(H-h)^2}}\begin{pmatrix}x\\ H-h\end{pmatrix} + \frac{T_2}{\sqrt{(L-l)^2+(H-h)^2}}\begin{pmatrix}L-l\\ H-h\end{pmatrix} = -W\begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}\,. \tag{3.9}$$

The astute observer may notice that the new force balance can be obtained by simply rotating the direction of gravity through an angle θ.

If, for some reason, we wish to express the tension in horizontal and vertical directions (the old frame of reference) then we must determine the transformation of the components from the new frame to the old one. Both tension are written in the form $\mathbf{T}_a = T\mathbf{t}_a = T\alpha\mathbf{w}_a$. Each $\mathbf{w}_a$ is a difference of two position vectors, for example, $(\mathbf{w}_1)_a = \mathbf{b}_a - \mathbf{e}_a$, where we have now made clear that these vectors are given in the new frame. By multiplying with the matrix $\mathcal{T}_{a\to e}$, we obtain

$$\mathcal{T}_{a\to e}(\mathbf{T}_1)_a = \mathcal{T}_{a\to e}\big(T_1\alpha_1(\mathbf{b}_a - \mathbf{e}_a)\big) \tag{3.10a}$$
$$= T_1\alpha_1\big(\mathcal{T}_{a\to e}\mathbf{b}_a - \mathcal{T}_{a\to e}\mathbf{e}_a\big) \tag{3.10b}$$
$$= T_1\alpha_1\big(\mathbf{b}_e - \mathbf{e}_e\big) \tag{3.10c}$$
$$= (\mathbf{T}_1)_e\,, \tag{3.10d}$$

which is the result we desire.

From the steps in (3.10), we observe that any vector can have its components transformed from one frame to another by multiplying with the transformation matrix $\mathcal{T}_{a\to e}$. This one matrix does the tranformation for all vectors. The key step that ensures this result is the one from (3.10a) to (3.10b) where we used property (2.28a) of matrix-vector multiplication. This property is so desirable that we make it the cornerstone of the definition of a linear transformation.

3.1 Definition

For a general definition we need a more abstract formulation of the matrix-vector multiplication. We write

$$\mathbf{y} = \mathcal{L}\{\mathbf{x}\}\,, \tag{3.11}$$

where $\mathcal{L}$ symbolizes an action that is performed on $\mathbf{x}$ to produce $\mathbf{y}$. We allow $\mathbf{x}$ and $\mathbf{y}$ to belong to any vector space, not necessarily the same ones. For matrix-vector multiplication

$$\mathcal{L}\{\mathbf{x}\} = A\mathbf{x}\,. \tag{3.12}$$

Now we restate (2.28a) as

$$\mathcal{L}\{\alpha\mathbf{x} + \beta\mathbf{y}\} = \alpha\mathcal{L}\{\mathbf{x}\} + \beta\mathcal{L}\{\mathbf{y}\}\,, \tag{3.13}$$

which must be true for all scalars α and β, and $\mathbf{x}$ and $\mathbf{y}$ must belong to the input vector space of $\mathcal{L}$. With this property, $\mathcal{L}$ is called a linear transformation (operator).

Surprisingly, and fortunately, there are many examples of linear operators that are useful in mathematical applications to science and engineering. Only two specific examples will be described here, but the expectation is that many more will arise in the natural course of study of mathematics and its applications.

3.2 The Transpose

To motivate the value of the transpose, let us consider the calculation of the total revenue generated by the sale of fruit baskets. Suppose the sales prices of the small, medium and large fruit baskets are \$0.75, \$1.50 and \$2.50

respectively. If we sell α_1 small fruit baskets, α_2 medium fruit baskets and α_3 large fruit baskets, then the revenue R will be

$$R = 0.75\alpha_1 + 1.50\alpha_2 + 2.50\alpha_3 \,. \tag{3.14}$$

As before, we may use a column vector to record the sales of fruit baskets

$$\mathbf{p} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \,. \tag{3.15a}$$

It seems useful to record the sales prices as a column vector as well

$$\mathbf{s} = \begin{pmatrix} 0.75 \\ 1.50 \\ 2.50 \end{pmatrix} \,. \tag{3.15b}$$

But how can we combine the vectors $\mathbf{p}$ and $\mathbf{s}$ to produce R?

Notice that (3.14) looks exactly the same as the first row of matrix-vector multiplication (2.25c). In other words,

$$R = \begin{bmatrix} 0.75 & 1.50 & 2.50 \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \,. \tag{3.16}$$

Can we have a matrix with only one row? Why not! That means we can have a matrix with only one column; a column vector is also just a special matrix. It seems, then, we should not list the prices of the fruit baskets as a column vector but rather as a row vector. But that is just too annoying: some vectors are rows, some are columns. Who wants to keep track? Instead, let's introduce the *transpose* operator that takes a column vector and rewrites it as a row vector. As an example, if

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \,, \quad \text{then} \quad \mathcal{L}\{\mathbf{x}\} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \,. \tag{3.17a}$$

Is this operator linear? Consider

$$\alpha\mathbf{x} + \beta\mathbf{y} = \begin{pmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \\ \alpha x_3 + \beta y_3 \end{pmatrix} \,. \tag{3.17b}$$

Then,

$$\mathcal{L}\{\alpha\mathbf{x} + \beta\mathbf{y}\} = \begin{bmatrix}(\alpha x_1 + \beta y_1) & (\alpha x_2 + \beta x_2) & (\alpha x_3 + \beta y_3)\end{bmatrix} . \tag{3.17c}$$

Since

$$\begin{aligned}\alpha\mathcal{L}\{\mathbf{x}\} + \beta\mathcal{L}\{\mathbf{y}\} &= \alpha \begin{bmatrix}x_1 & x_2 & x_3\end{bmatrix} + \beta \begin{bmatrix}y_1 & y_2 & y_3\end{bmatrix} \\ &= \begin{bmatrix}(\alpha x_1 + \beta y_1) & (\alpha x_2 + \beta x_2) & (\alpha x_3 + \beta y_3)\end{bmatrix} ,\end{aligned} \tag{3.17d}$$

which agrees with (3.17c), and the operator is linear.

Obviously we should name this operator and one obvious possiblility is to write it as $\mathcal{L}\{\mathbf{x}\} = \mathcal{T}\{\mathbf{x}\}$. However, as is so common in mathematics, the standard approach is to introduce a symbolic pattern that express the operation: we use a superscript T. So,

$$\mathbf{x}^T = \begin{pmatrix}x_1 \\ x_2 \\ x_3\end{pmatrix}^T = \begin{bmatrix}x_1 & x_2 & x_3\end{bmatrix} . \tag{3.18a}$$

With this new symbol, The statement that the operation is linear is

$$(\alpha\mathbf{x} + \beta\mathbf{y})^T = \alpha\mathbf{x}^T + \beta\mathbf{y}^T . \tag{3.18b}$$

And, finally, we write (3.14) as $R = \mathbf{s}^T\mathbf{p}$.

3.3 Operators in function spaces

Now let us consider completely different vector spaces, the spaces P_n of polynomials of degree $n-1$ that is introduced in (1.27). What can we do with polynomials? Why, we can differentiate them! We interpret differentiate as an operator acting on a polynomial (vector) in P_n to produce another polynomial of one degree less, i.e., a polynomial (vector) in P_{n-1}. Specifically, let $\mathbf{p} \in P_4$ be written as

$$\mathbf{p} = a_0 + a_1 x + a_2 x^2 + a_3 x^3 . \tag{3.19}$$

Introduce $\mathcal{D}\{\mathbf{p}\}$ as the differentiation operator that produces $\mathbf{q} \in P_3$,

$$\begin{aligned}\mathbf{q} = \mathcal{D}\{\mathbf{p}\} &\equiv \frac{d}{dx}\left(a_0 + a_1 x + a_2 x^2 a_3 x^3\right) \\ &= a_1 + 2a_2 x + 3a_3 x^2 .\end{aligned} \tag{3.20}$$

Is this operator linear? Consider

$$\begin{aligned}\mathcal{D}\{\alpha\mathbf{v}+\beta\mathbf{w}\} &= \mathcal{D}\{(\alpha v_0+\beta w_0)+(\alpha v_1+\beta w_1)x \\ &\qquad\qquad + (\alpha v_2+\beta w_2)x^2+(\alpha v_3+\beta w_3)x^3 \\ &= (\alpha v_1+\beta w_1)+2(\alpha v_2+\beta w_2)x+3(\alpha v_3+\beta w_3)x^2\,, \end{aligned} \tag{3.21a}$$

and

$$\begin{aligned}\alpha\mathcal{D}\{\mathbf{v}\}+\beta\mathcal{D}\{\mathbf{w}\} &= \alpha(v_1+2v_2x+3v_3x^2)+\beta(w_1+2w_2x+3w_3x^2) \\ &= (\alpha v_1+\beta w_1)+2(\alpha v_2+\beta w_2)x+3(\alpha v_3+\beta w_3)x^2\,. \end{aligned} \tag{3.21b}$$

Since the results agree, the operator is linear.

The real value in this example is that it demonstrates how we may convert operators on vector spaces into matrices and column vectors. This is how. First, we write $\mathbf{p}$ as the member of a span of basis polynomials (2.11):

$$\mathbf{p} = a_0\mathbf{e}_1 + a_1\mathbf{e}_2 + a_2\mathbf{e}_3 + a_3\mathbf{e}_4\,, \tag{3.22a}$$

where

$$\mathbf{e}_1 \equiv 1\,,\quad \mathbf{e}_2 \equiv x\,,\quad \mathbf{e}_3 \equiv x^2\,,\quad \mathbf{e}_4 \equiv x^3\,. \tag{3.22b}$$

Second, we apply the operator to the basis polynomials:

$$\begin{aligned} &\mathbf{d}_1 = \mathcal{D}\{\mathbf{e}_1\} = 0\,, &&\mathbf{d}_2 = \mathcal{D}\{\mathbf{e}_2\} = 1 = \mathbf{e}_1\,, \\ &\mathbf{d}_3 = \mathcal{D}\{\mathbf{e}_3\} = 2x = 2\mathbf{e}_2\,, &&\mathbf{d}_4 = \mathcal{D}\{\mathbf{e}_4\} = 3x^2 = 3\mathbf{e}_3\,. \end{aligned} \tag{3.22c}$$

Third, we use the fact that the operator is linear:

$$\begin{aligned}\mathcal{D}\{\mathbf{p}\} &= a_0\mathcal{D}\{\mathbf{e}_1\}+a_1\mathcal{D}\{\mathbf{e}_2\}+a_2\mathcal{D}\{\mathbf{e}_3\}+a_3\mathcal{D}\{\mathbf{e}_4\} \\ &= a_0\mathbf{d}_1+a_1\mathbf{d}_2+a_2\mathbf{d}_3+a_3\mathbf{d}_4\,. \end{aligned} \tag{3.22d}$$

Fourth, we convert the results into column vectors:

$$\mathbf{p} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}\,,\quad \mathbf{d}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\,,\quad \mathbf{d}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\,,\quad \mathbf{d}_3 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}\,,\quad \mathbf{d}_4 = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}\,. \tag{3.22e}$$

Note that $\mathbf{q} \in \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ so there are only three components for each $\mathbf{d}_j$. Fifth, we interpret (3.22d) as a matrix-vector multiplication where the columns of the matrix are $\mathbf{d}_j$:

$$\begin{aligned}\mathbf{q} &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \\ &= \begin{pmatrix} a_1 \\ 2a_2 \\ 3a_3 \end{pmatrix} \end{aligned} \tag{3.22f}$$

Hopefully the analogy with (2.24) and (3.3) is clear.

4 Summary

The most useful perspective on vector spaces is that they are specified as spans. This perspective allows us to take the coefficents of a linear combination and treat them as though they are a column vectors and column vectors are natural well-suited for computers. Further, linear operations with vectors can be cast as matrix-vector multiplications, another action well-suited for computers.

What we have not fully exploited yet are the ideas associated with angles and distances. The inclusion of the appropriate generalizations of these ideas with vector spaces and linear operators opens up even more power in the applications of matrices and vectors.