Basics of Linear Algebra

John Leo

October 9, 2010

1 Notation

Mathematics is a language with its own vocabulary and abbreviations. Just as you write LOL and ROFL for common expressions so you don't have to write out all the words, mathematicians use abbreviations for common objects and expressions to save time. This is especially valuable in class, both when writing on the board and taking notes, so I'll use these abbreviations often, and I encourage you to use them in your own presentations and notes as well.

First of all numbers are very important in math, and there are standard abbreviations for the most common types of numbers we will talk about:
$\mathbb{N}$: The set of natural numbers, $\{1, 2, 3, \dots\}$. Notice that we write a set using curly brackets.
$\mathbb{Z}$: The set of integers, $\{\dots, -2, -1, 0, 1, 2, \dots\}$. The letter $\mathbb{Z}$ comes from the German word Zahl, meaning number.
$\mathbb{Q}$: The set of rational numbers. Here $\mathbb{Q}$ comes from the word quotient. Precisely this is $\left\{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\right\}$. This notation looks fancy and we'll explain it shortly, but in English this can be read "the set of fractions $\frac{m}{n}$ such that m and n are integers and $n \neq 0$."
$\mathbb{R}$: The real numbers. These can be defined precisely but it is very tricky to do so, so I'll just assume you basically know what these are. They are the rational numbers plus the irrational numbers (the latter being numbers with non-repeating decimals).
$\mathbb{C}$: The complex numbers. This is $\{a + bi : a, b \in \mathbb{R}\}$, where $i = \sqrt{-1}$. We'll be studying these later in the course.

I've already introduced some other notation which I'll now explain. A set is indicated by putting curly brackets around all of the members. The sets of numbers above were all infinite, but there are finite sets also. For example $\{1, 2, 3\}$ is a set with three elements: 1, 2 and 3. If you want to put a condition on what is in the set, you use a colon, which is read as "such that". So $\{n : n \in \mathbb{Z}, n \geq 0\}$ is read "the set of numbers n such that n is an integer and $n \geq 0$", meaning the set $\{0, 1, 2, 3, \dots\}$ which is sometimes called the "whole numbers" (I remember this because 0 has a hole). Notice the symbol $\in$, which means "in" or "is an element of" (it kind of looks like an E).

Another abbreviation for "such that" that I often use when writing on the board is "s.t.". Two other common symbols are $\forall$, which means "for all", and $\exists$ which means "exists". So the mathematical statement

$$\forall x \in \mathbb{R}, \exists y \in \mathbb{R} \text{ s.t. } x + y = 0$$

is read in English as "for all real numbers x there exists a real number y such that $x+y = 0$", which is the statement that additive inverses exist.

Don't let the notation scare you! Think of math as a foreign language, but a very easy one that translates directly into English!

2 Vectors

In this class we distinguish *points* and *vectors*. A *point* is just a point in space. Let's say our space is two-dimensional to be concrete. The fancy name for two dimensional space is $\mathbb{R}^2$, where $\mathbb{R}$ denotes the set of real numbers. This indicates that every point can be described as a pair of real numbers, once we pick a coordinate system. The point $(0, 0)$ is the origin (center) of the coordinate system. The pair $(3, 2)$ refers to the point that is 3 units to the right and 2 units up from the origin.

A *vector*, on the other hand, indicates a displacement. So the vector $[3, 2]$ means "move 3 units to the right and 2 units up". This sounds a lot like how we just described the point $(3, 2)$, but notice the difference: With the point, we had to start at the origin. With a vector, you can start anywhere. More subtly, the two represent different concepts. A point represents a position in space, whereas a vector represents a displacement—the act of moving from one position to another. If this seems confusing don't feel bad—it is confusing!

We draw a vector as an arrow to graphically indicate the displacement. A vector is also characterized completely by its length and direction, which is another good way to think about it, but for our purposes it is better to think of it as a displacement.

A key difference between points and vectors is that you can add vectors but you can't add points. It isn't at all clear geometrically what the "sum" of two points would mean. On the other hand it's very clear what the sum of two displacements would be. If we want to add $[3, 2]$ and $[-4, 5]$ it means we should go 3 units to the right and 2 up, and then follow that by moving 4 units to the left and 5 up. That's the same as moving 3 units to the right followed by 4 units to the left (a net change of 1 unit to the left), and then moving 2 units up and then 5 units up (a net change of 7 units up). So we can add vectors by adding their components. This is expressed simply as $[3, 2] + [-4, 5] = [-1, 7]$.

You can also multiply a vector by a real number, which is called a *scalar*. This operation is then called *scalar multiplication*. So $5[3, 2]$ means to do 5 displacements of 3 to the right and 2 up. It's clear that $5[3, 2] = [15, 10]$. We are "scaling" the vector by 5.

One other neat thing you can do is add a vector to a point! That makes sense geometrically as well. It means you start at the point, then follow a displacement, and end up at another point. For example $(3, 2) + [6, -7] = (9, -5)$.

We have already seen three-dimensional points and vectors as well. The space in this case is called $\mathbb{R}^3$. The points give positions (x, y, z) and the vectors are now three-dimensional displacements. For example $[3, 2, -4]$ means go 3 in the positive x direction, 2 in the positive y direction, and 4 in the negative z direction. Again vectors are written with arrows, and can be added both to vectors and to points, and multiplied by scalars. It's easy to generalize to n-dimensional space, $\mathbb{R}^n$. Think about what $\mathbb{R}^1$ and $\mathbb{R}^0$ would mean!

Note that vectors are often written in column form, like this:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

but we will usually write them as row vectors $[x, y, z]$ here to save space. As is conventional, we use bold letters near v (such as $\mathbf{u}, \mathbf{v}, \mathbf{w}$) to refer to vectors, and letters either near the beginning (like a, b, c) or end (such as x, y, z) of the alphabet to refer to scalars. On the blackboard, vectors are written with arrows, such as $\vec{v}$.

Exercises

1. What exactly is the vector space $\mathbb{R}^1$? What do you think $\mathbb{R}^0$ should mean?

3 Vector Spaces

Let's forget about points for the time being, and concentrate only on vectors. Let's consider the vectors in $\mathbb{R}^2$, which look like $\mathbf{v} = [a, b]$, where $a, b \in \mathbb{R}$. Let $\mathbf{w} = [c, d]$, where $c, d \in \mathbb{R}$, and let $e \in \mathbb{R}$ be a scalar. Then vector addition, as we noted in the previous section, is defined as $\mathbf{v} + \mathbf{w} = [a, b] + [c, d] = [a + c, b + d]$. Notice that it doesn't matter which order we add the vectors in: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

Scalar multiplication is defined as $e\mathbf{v} = e[a, b] = [ea, eb]$.

There is a "zero" vector $\mathbf{0} = [0, 0]$ which when added to any vector leaves the vector unchanged. So $\mathbf{v} + \mathbf{0} = [a, b] + [0, 0] = [a, b] = \mathbf{v}$. As a displacement, $\mathbf{0}$ means don't move at all.

Also for any vector there is an inverse, a sort of "opposite displacement" that says go the same distance in the opposite direction, so that the two combined give zero displacement. It's clear that if $\mathbf{v} = [a, b]$ then $-\mathbf{v} = [-a, -b]$ and that $\mathbf{v} + (-\mathbf{v}) = [a, b] + [-a, -b] = [0, 0] = \mathbf{0}$.

You can see that vectors satisfy properties similar to those for ordinary numbers that you learned back in middle school. There is in fact a complete list of the fundamental properties that vector addition and scalar multiplication satisfy. Compare them to the ones you know for real numbers.

For all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and all scalars a, b:

1. Associative law: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
2. Additive identity: There exists a unique vector $\mathbf{0}$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
3. Additive inverse: There exists a unique vector, denoted $-\mathbf{v}$, such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
4. Commutative law: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
5. Scalar multiplication identity: $1\mathbf{v} = \mathbf{v}$.
6. Scalar multiplication associativity: $(ab)\mathbf{v} = a(b\mathbf{v})$.
7. Distributive law (vector addition): $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
8. Distributive law (scalar addition): $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$.

Something that satisfies these properties is known as a *vector space*. So $\mathbb{R}^2$ and $\mathbb{R}^3$ and in fact $\mathbb{R}^n$ for any $n \geq 0$ is a vector space. However it turns out there are a lot of other vector spaces around which satisfy these properties. We're not going to study them this year, but you'll run across them in the not too distant future, and understand better what a powerful concept a vector space is. For now I just want you to realize the concept exists.

Exercises

1. Show that in $\mathbb{R}^2$, with vector addition and scalar multiplication defined as usual, that the vector $[0, 0]$ is the unique additive identity.
2. Given vector $[x, y] \in \mathbb{R}^2$, what is its additive inverse? Prove that this inverse is unique.

4 Dot Product, Length and Angle

In a vector space we can add two vectors together, and multiply a vector by a scalar, but there is no operation to multiply two vectors together and get another vector. Indeed in $\mathbb{R}^2$ it's not clear what it would even mean to "multiply" two vectors together.

It is, however, possible to define a sort of product of vectors that produces a scalar. This is known as the *dot product* in $\mathbb{R}^n$, and it is an example of a more general type of product known as an *inner product*. We will denote the dot product with $\cdot$ just as the name implies. The dot product is defined for $\mathbb{R}^2$ as $[a, b] \cdot [x, y] = ax + by$ and for $\mathbb{R}^3$ as $[a, b, c] \cdot [x, y, z] = ax + by + cz$ (see Exeter 9:7). You should see how to generalize this to $\mathbb{R}^n$. What is the dot product for $\mathbb{R}^1$?

The dot product we have defined satisfies these general properties of an inner product, which hold for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a \in \mathbb{R}$.

1. Linearity (vector addition): $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$.
2. Linearity (scalar multiplication): $a(\mathbf{u} \cdot \mathbf{v}) = (a\mathbf{u}) \cdot \mathbf{v}$.
3. Commutativity: $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.
4. Positive definiteness: $\mathbf{u} \cdot \mathbf{u} > 0$ whenever $\mathbf{u} \neq \mathbf{0}$.

You have actually already shown that the dot product satisfies some of these properties (for $\mathbb{R}^3$) in Exeter 10:11, and you'll have a chance to show the others in the exercises below. The first two properties show that the dot product is "linear" in its first argument; we will discuss linearity in more depth in the next section. It is also linear in the second argument using commutativity.

The importance of the dot product, and inner product in general, is that it relates to the geometric concepts of length and angle. Recall that the *length* (also called *magnitude*) of vector $[a, b] \in \mathbb{R}^2$ is $\sqrt{a^2 + b^2}$ by the Pythagorean theorem. Similarly we can use two applications of the Pythagorean theorem to show that the length of $[a, b, c] \in \mathbb{R}^3$ is $\sqrt{a^2 + b^2 + c^2}$, and the obvious generalization works for $\mathbb{R}^n$ (it's hard to visualize for $n > 3$, but with some thought you may be able to understand it conceptually). Note that for $\mathbb{R}^1$ the length of $[a]$ is $\sqrt{a^2}$, which is just another way to say $|a|$, since $\sqrt{a^2}$ refers to the positive square root of a^2.

The length of vector $\mathbf{v} \in \mathbb{R}^n$ is denoted $|\mathbf{v}|$ and by definition of both length and dot product we have $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$, or equivalently (since everything is non-negative) $|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$. So dot product can be used to determine length.

More surprising is that dot product can be used to determine angle as well! In Exeter 15:3 you show that

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}|\,|\mathbf{v}|} \tag{1}$$

where θ is the angle between vectors $\mathbf{u}$ and $\mathbf{v}$. Since we can express the lengths of $\mathbf{u}$ and $\mathbf{v}$ using dot product, this means that $\cos\theta$ is determined completely from the dot product, and θ essentially determined as well.

The advantage of this approach is that an inner product is a more general concept than length and angle, and it is possible to define an inner product even when it is not clear geometrically how one would define length and angle, and then define length and angle using the inner product.

Exercises

1. Show that the dot product for $\mathbb{R}^3$ is positive definite.

2. In Exeter 10:11(d), you showed that $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ in $\mathbb{R}^3$. Use this to prove the linearity (vector addition) law above, solely using three applications of the commutative law.

3. Prove linearity (scalar multiplication) for $\mathbb{R}^3$.

4. Exeter 15:3.

5. Exeter 14:10.

6. Exeter 15:6.

7. Exeter 17:6.

8. Exeter 16:9.

9. Exeter 18:3.

10. Given $\cos\theta$ what are the possible values of θ? In $\mathbb{R}^n$ is it possible geometrically to narrow down the choice to one? What would this imply about defining an inverse to cosine (called $\cos^{-1}$ or arccos)?

11. What are the minimum and maximum values of $\cos\theta$? Notice that this implies we had better have
$$|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}| \cdot |\mathbf{v}|$$
if Equation 1 is going to make sense. (Notice that the absolute value on the left is normal absolute value, not length, and the dot on the right refers to normal multiplication, not dot product!) Fortunately this is true for any inner product that satisfies the properties given above, and is the famous *Cauchy-Schwarz Inequality*. Prove this for $\mathbb{R}^2$ using the dot product.

12. Exeter 18:10.

13. Exeter 19:9.

14. Exeter 20:4.

15. Exeter 21:7.

16. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ be four vectors all with length 1 and with angle (with respect to the postive x-axis) of $0, \pi/2, \pi, 3\pi/2$ radians respectively. (Recall 2π radians is 360 degrees.) What is $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4$? What is $\mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4$?

17. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6$ be six vectors all with length 1 and with angle (with respect to the postive x-axis) of $0, \pi/3, 2\pi/3, \pi, 4\pi/3, 5\pi/3$, radians respectively. What is $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 + \mathbf{v}_5 + \mathbf{v}_6$? What is $\mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 + \mathbf{v}_5 + \mathbf{v}_6$?

18. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5$ be five vectors all with length 1 and with angle (with respect to the postive x-axis) of $0, 2\pi/5, 4\pi/5, 6\pi/5, 8\pi/5$, radians respectively. What is $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 + \mathbf{v}_5$? What is $\mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 + \mathbf{v}_5$?

5 Linear Transformations

We have seen *transformations* appear in several exercises so far (for example 6:6, 7:1, 7:11, 9:1) and for those who used the Exeter *Mathematics 2* book last year you should already be very familiar with transformations. The *Mathematics 2* reference section defines a transformation to be "a function that maps points to points." So far they've assumed these points to be in $\mathbb{R}^2$, and so the transformation looks something like $T(x, y) = (x', y')$, where (x', y') is the *image*, or result, of transforming point (x, y).

We now come to something confusing. Tranformations map points to points, but we are now going to think of the points as vectors! This allows us to use all of the operations and properties of vector spaces we have developed. We can think of the points as vectors by going back to the idea that a point is just a "displacement" from the origin. So $(3, 2)$ as a point means go over 3 and up 2 from the origin, and we think of it as the vector $[3, 2]$ which is forced to have its tail at the origin, and its head at $(3, 2)$. Now we are allowed to "add" two points by thinking of them as vectors, but the resulting "point" must be a vector whose tail is at the origin. So we think of $(3, 2)$ and $(-4, 5)$ as vectors $[3, 2]$ and $[-4, 5]$, and when we add them we get the vector $[-1, 7]$, which we think of having its tail at $(0, 0)$ and head at $(-1, 7)$, and so it corresponds to the point $(-1, 7)$.

In this way we get something called a "one-to-one correspondence" between points and vectors: for every point there is a vector, and for every vector there is a point. The correspondence should be clear: the point (x, y) corresponds to the vector $[x, y]$, and vice-versa. This is also called a *bijection* (which you may prefer if you know French). Exeter starts to treat points as vectors around now in the text, but they never explain what they are doing, so it is very confusing.

There are many different types of transformations in *Mathematics 2* such as translations, rotations, dilations, reflections and so forth. There are also classes of transformation such as *isometries*, which preserve distance between points, and *similarity transformations*, in which all lengths may be multiplied by the same scalar. We are now starting to reconsider these in the context of vectors.

In vector notation, we can simply write $T(\mathbf{u}) = \mathbf{v}$, where $\mathbf{v}$ is the image of vector $\mathbf{u}$ under transformation T. It is typical to use capital letters near T for transformations. As an example, if T is dilation about the origin with magnification factor 3, then it is written using points as $T(x, y) = (3x, 3y)$, and with vectors simply as $T(\mathbf{v}) = 3\mathbf{v}$.

An important class of transformation that Exeter does not describe explicitly but which is fundamental in everything we are doing is the *linear transformation*. A linear transformation is one which preserves the operations of adding two vectors and multiplying by a scalar. So a linear transformation satisfies the following two properties for all vectors $\mathbf{u}$ and $\mathbf{v}$ and all scalars a:

1. Linearity (vector addition): $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$.

2. Linearity (scalar multiplication): $T(a\mathbf{v}) = aT(\mathbf{v})$.

Notice how these are similar to the two linearity properties of the dot product.

As an example, the dilation transformation $T(\mathbf{v}) = 3\mathbf{v}$ is linear because $T(\mathbf{u} + \mathbf{v}) = 3(\mathbf{u} + \mathbf{v}) = 3\mathbf{u} + 3\mathbf{v} = T(\mathbf{u}) + T(\mathbf{v})$, and also $T(a\mathbf{v}) = 3a\mathbf{v} = a(3\mathbf{v}) = aT(\mathbf{v})$.

Note that linearity is saying that the transformation "commutes" in a certain sense with vector addition and scalar multiplication. That is, it doesn't matter which order we do things in. We can add two vectors first and then perform the transformation, or else we can transform the two vectors separately and then add them—in either case the result is the same. And the same goes for scalar multiplication: We can multiply by a scalar first and then perform the transformation, or perform the transformation first and then multiply by the scalar, and the result is the same. Functions such as linear transformations which preserve some specific structure (in this case vector addition and scalar multiplication) of some mathematical object (in this case vectors) are of fundamental importance to mathematics and you will see them again and again as you learn more.

Exercises

1. Exeter 16:6.

2. Exeter 16:7.

3. Which of the transformations of Exeter 7:1, 7:11 and 9:1 are linear? Either prove that each is linear or demonstrate a counterexample (either a pair of vectors or a scalar/vector pair that fail to transform linearly).

4. Which of the transformations of Exeter 16:6 are linear? Again either prove linearity or find a counterexample.

5. Consider the translation $T(x, y) = (x + 1, y + 1)$. In vector notation, $T(\mathbf{v}) = \mathbf{v} + [1, 1]$. Is T linear? Again either prove linearity or find a counterexample.

6. We say that transformation T is an isometry if $|T(\mathbf{v})| = |\mathbf{v}|$ for all $\mathbf{v}$ in the vector space. Say that our vector space is $\mathbb{R}^2$. Recall from *Mathematics 2* that a transformation is an isometry if it preserves distances between any two points. In the definition here, it seems that we are only saying that the transformation preserves the distance between points and the origin (the length of the vectors). Why is this sufficient to show that in fact the distance between any two points is preserved?

6 Basis and Dimension

A *basis* for a vector space is a set of vectors that can be used to generate all of the vectors in a unique way. For example, in $\mathbb{R}^2$ the vectors $[1, 0]$ and $[0, 1]$ generate all of the other vectors as follows. If we are given an arbitrary vector $[x, y]$ in $\mathbb{R}^2$, we can write it as $[x, y] = x[1, 0] + y[0, 1]$. We say that $[x, y]$ is a *linear combination* of $[1, 0]$ and $[0, 1]$, since we can write $[x, y]$ as a sum of the two vectors $[1, 0]$ and $[0, 1]$, with appropriate scalar coefficients. Furthermore this representation of $[x, y]$ is unique, because if also $[x, y] = a[1, 0] + b[0, 1]$, then since $a[1, 0] + b[0, 1] = [a, b]$ we must have $a = x$ and $b = y$, so x and y were the only coefficients that worked.

The vectors $[1, 0]$ and $[0, 1]$ are known as the *standard basis* for $\mathbb{R}^2$, and are often designated as $\mathbf{e}_1$ and $\mathbf{e}_2$ respectively. It is easy to see how to generalize this to $\mathbb{R}^n$. For example the standard basis for $\mathbb{R}^3$ is $\mathbf{e}_1 = [1, 0, 0]$, $\mathbf{e}_2 = [0, 1, 0]$ and $\mathbf{e}_3 = [0, 0, 1]$. In applied math and physics books you will sometimes see these three vectors written as $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ or $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$. The carat or "hat" over the vector indicates that it has length 1.

There can be many different bases for a vector space. To determine if a set of vectors is a basis, you need to make sure every vector in the vector space can be written as a linear combination of the basis vectors in exactly one way. It is this uniqueness that makes a basis so powerful.

Another basis for $\mathbb{R}^2$ is $[1,0]$ and $[1,1]$. Let's prove this.

Theorem. *The vectors* $[1,0]$ *and* $[1,1]$ *are a basis for* $\mathbb{R}^2$*.*

Proof. We can write an arbitary vector $[x,y]$ as

$$(x-y)[1,0]+y[1,1]=[x-y,0]+[y,y]=[x,y].$$

Furthermore say that $[x,y]=a[1,0]+b[1,1]$ for some a and b. Then since $a[1,0]+b[1,1]=[a+b,b]$, then we must have $b=y$ and $a+b=x$, and so $a=x-b=x-y$. Therefore the representation $[x,y]=(x-y)[1,0]+y[1,1]$ is unique. Since every vector in $\mathbb{R}^2$ can be written as a unique linear combination of $[1,0]$ and $[1,1]$, these two vectors form a basis for $\mathbb{R}^2$. □

Although there may be many different bases for a vector space, one remarkable fact is that they always have the same number of elements! For example a basis for $\mathbb{R}^1$ must always consist of one vector (which vectors would work?), a basis for $\mathbb{R}^2$ must always consist of two vectors, a basis for $\mathbb{R}^3$ must always consist of three vectors, and so forth. We will not prove this fact here. The number of vectors in the basis for a vector space is called the *dimension* of the vector space. So $\mathbb{R}^1$ has one dimension, $\mathbb{R}^2$ has two dimensions, $\mathbb{R}^3$ has three dimensions and in general $\mathbb{R}^n$ has n dimensions, just as you would think. Geometrically a basis is capturing the notion of "independent degrees of freedom", with each basis element representing a distinct degree of freedom.

Exercises

1. Prove that $[1,1]$ and $[-1,1]$ form a basis for $\mathbb{R}^2$, imitating the style of the theorem above.

2. Do $[1,1]$ and $[-1,-1]$ form a basis for $\mathbb{R}^2$? If so, prove it. If not, show what goes wrong.

3. Prove that $[1,0,0]$, $[0,1,0]$ and $[1,0,1]$ form a basis for $\mathbb{R}^3$, imitating the style of the theorem above.

4. Why doesn't $[1,1]$ form a basis for $\mathbb{R}^2$? What fails?

5. Why don't the three vectors $[1,0]$, $[0,1]$ and $[1,1]$ form a basis for $\mathbb{R}^2$? What fails?

7 Linear Transformations and Bases

We are now ready to put together the ideas of linear transformations and bases. The amazing fact is that once we know how linear transformation T transforms all of the basis vectors, we also know how it transforms any vector! This follows immediately from the fact that we can write any vector as a unique linear combination of the basis vectors, and the fact that T is linear.

Let's see how this works with an example. Let's consider $\mathbb{R}^2$ with the standard basis. Then $[x,y]=x[1,0]+y[0,1]$ in terms of this basis. If we now apply T to $[x,y]$ we get

$$T([x,y])=T(x[1,0]+y[0,1])=T(x[1,0])+T(y[0,1])=xT([1,0])+yT([0,1]).$$

Notice that we used both linearity rules for T. And notice that given $T([1,0])$ and $T([0,1])$, the images of $[1,0]$ and $[0,1]$ respectively, we know exactly what $T([x,y])$ must be!

Let's see how to use this to our advantage. Say we want to dilate by 3 centered at the origin. All we need to do is figure out the images of $[1,0]$ and $[0,1]$ by this dilation, and then we can use linearity to find out the image of any vector. So $T([1,0]) = [3,0]$, $T([0,1]) = [0,3]$, and using the fact that $T([x,y]) = xT([1,0]) + yT([0,1])$ (that we showed above) we get in general that $T([x,y]) = x[3,0] + y[0,3] = [3x, 3y]$.

That was so simple perhaps it didn't seem interesting, so let's try a more complicated example. Consider a reflection across the line $y = x$. This was on the practice test and you may remember there were several ways to do it, all somewhat complicated. We will now see how easy it is to do with our basis. It is easy to see that $T([1,0]) = [0,1]$ and $T([0,1]) = [1,0]$ in this case. The two basis vectors are simply exchanged. Now it follows that in general $T([x,y]) = x[0,1] + y[1,0] = [y,x]$, exactly what we showed before but now it is a lot less work.

One more example, a rotation counterclockwise by 45 degrees: Here we have $T([1,0]) = \left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$ and $T([0,1]) = \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$, which you can calculate using 45-45-90 triangles. In this case calculating the tranformation of a general point $[x,y]$ would be very complicated, but using linearity it is simple:

$$T[x,y] = x\left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right] + y\left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right] = \left[\frac{\sqrt{2}}{2}(x-y), \frac{\sqrt{2}}{2}(x+y)\right]$$

There is one subtlety that needs to be addressed: This only works if T is really a linear transformation! Exercise 1 shows what can go wrong if that assumption is not satisfied.

Exercises

1. Consider the translation $T([x,y]) = [x,y] + [1,1] = [x+1, y+1]$. We have $T([1,0]) = [2,1]$ and $T([0,1]) = [1,2]$, so by linearity $T([x,y]) = x[2,1] + y[1,2] = [2x+y, x+2y]$, which isn't a translation by $[1,1]$ at all! What's going on?

2. Write the formula for $T([x,y])$ where T is rotation counterclockwise by θ degrees.

3. Try writing a formula for $T([x,y])$, where T is dilation about the origin with magnification factor 3, using $[1,0]$ and $[1,1]$ as the basis vectors. That is, determine what T does to the two basis vectors, and then compute what T does to an arbitrary vector by using how the vector is written as a linear combination of $[1,0]$ and $[1,1]$. Do you get the same answer as we got using $[1,0]$ and $[1,1]$ as the basis?

4. Write the formula for $T([x,y])$ where T is dilation about the origin with magnification factor 5, followed by reflection about the line $y = x$.

5. Write the formula for $T([x,y])$ where T is reflection about the line $y = x$, followed by counterclockwise rotation by 30 degrees, followed by dilation about the origin with magnification factor 7.

8 Matrices

Once a basis is fixed for a vector space, it is possible to represent a linear transformation in a very concrete form using a matrix. Let's continue to use $\mathbb{R}^2$ with the standard basis to be concrete. We are now going to write our vectors as column vectors, which is standard when using matrices. Say that

$$T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}a\\c\end{bmatrix} \text{ and } T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}b\\d\end{bmatrix}.$$

Then by linearity for an arbitary vector $[x, y]$ we have

$$\begin{aligned} T\left(\begin{bmatrix}x\\y\end{bmatrix}\right) &= x\,T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) + y\,T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) \\ &= x\begin{bmatrix}a\\c\end{bmatrix} + y\begin{bmatrix}b\\d\end{bmatrix} \\ &= \begin{bmatrix}ax+by\\cx+dy\end{bmatrix}. \end{aligned}$$

Now note that the first entry of the result vector is just $[a, b] \cdot [x, y]$ and the second entry is just$[c, d] \cdot [x, y]$. This is exactly how we defined matrix multiplication! So we can write

$$T\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{bmatrix}a & b\\c & d\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}.$$

This gives us a very concrete representation of the linear transformation as matrix multiplication. Note also that the two columns of the matrix are exactly the images of the basis vectors $[1, 0]$ and $[0, 1]$ under the transformation. This gives an easy way to write the matrix for any linear transformation. The Exeter exercises explore this further.

Exercises

1. Exeter 16:8.
2. Exeter 18:1.
3. Exeter 18:8.
4. Exeter 18:9.
5. Exeter 19:3.
6. Exeter 19:4.
7. Exeter 20:7.
8. Exeter 21:4.
9. Exeter 21:6.
10. Exeter 22:1.
11. Exeter 22:3.
12. Exeter 22:6.
13. Exeter 22:7.