Vectors and vector spaces

Vectors are introduced in Multivariate Calculus and beginning Physics courses in seemingly different ways. Introductory physics textbooks favor the geometric definition: vectors are directed line segments which can be added using the Parallelogram Law, shown in Figure 1:

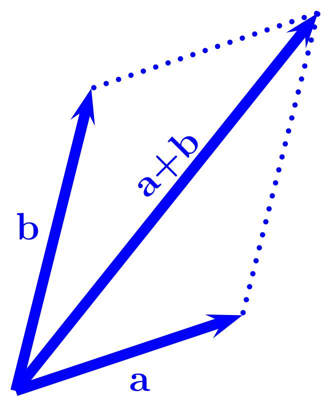

Figure 1: Parallelogram Law

From the Parallelogram Law follows that adding a vector with itself produces a vector with the same direction but twice the length. This suggests a scaling operation: if s is a positive scalar and $\mathbf{v}$ is a vector, the scalar product $s\,\mathbf{v}$ is the vector in the same direction as $\mathbf{v}$ but with s-times the length; if s is negative then upon scaling the length of $\mathbf{v}$ by $|s|$ one also reverses the direction.

In Multivariate Calculus vectors are *motivated* using geometry and physics, however, the "official" Calculus definition is stated in purely algebraic terms. A Calculus vector is an n-tuple of numbers such as $\langle a, b, c\rangle$. What distinguishes a vector $\langle a, b, c\rangle$ from a point (a, b, c) is the rule of addition

$$\langle a_1, b_1, c_1\rangle + \langle a_2, b_2, c_2\rangle = \langle a_1 + a_2, b_1 + b_2, c_1 + c_2\rangle \qquad (1)$$

which is consistent with Figure 1. (Note that points in space look a lot like vectors, yet it does not make sense to add points coordinate-wise. Think, for example, of adding the coordinates of Stockton and San Francisco: the result has no geographic meaning.) The scaling operation in Calculus becomes

$$s\,\langle a, b, c\rangle = \langle s\,a, s\,b, s\,c\rangle \tag{2}$$

which, again, is consistent with the Parallelogram Law.

In Calculus one also learns that vectors can be “multiplied” in two different ways. The dot-product

$$\langle a_1, b_1, c_1\rangle \cdot \langle a_2, b_2, c_2\rangle = a_1\,a_2 + b_1\,b_2 + c_1\,c_2$$

is useful for measuring lengths and angles owing to the formula:

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\,\cos(\theta).$$

Here we use standard notation $|\mathbf{a}|$ for the length of $\mathbf{a}$ and denote by θ the angle between $\mathbf{a}$ and $\mathbf{b}$. Also, in three dimensions, vectors can be crossed:

$$\langle a_1, b_1, c_1\rangle \times \langle a_2, b_2, c_2\rangle = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

The definition of the cross-product necessitates the introduction of 3-by-3 determinants. These are defined in terms of 2-by-2 determinants as shown below:

$$\begin{aligned} \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}\mathbf{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}\mathbf{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\mathbf{k} \\ &= (a_2\,b_3 - a_3\,b_2)\,\mathbf{i} - (a_1\,b_3 - a_3\,b_1)\,\mathbf{j} + (a_1\,b_2 - a_2\,b_1)\,\mathbf{k}. \end{aligned}$$

Unfortunately, introducing vectors as arrows but treating them as n-tuples of numbers that can be manipulated using disconnected numerical recipes leads to cognitive difficulties which we discuss below.

What are vectors, really?

When this question is first asked in any introductory Linear Algebra course the typical answers include:

1. Magnitude and direction.

2. Arrows.
3. Numbers within angles, e.g., $\langle 3, 5, 1 \rangle$.
4. Expressions like $3\,\mathbf{i} + 5\,\mathbf{j} + \mathbf{k}$.

Note that all of the above answers are purely descriptive in nature and can be paraphrased: "A vector is whatever looks like…" There are two interrelated problems with descriptive definition of vectors. Firstly, it is not very clear what can be done with vectors, which vector operations are important and which are secondary. Secondly, it is very difficult to treat objects as vectors when the objects do not look like vectors, yet this is the core philosophy of Linear Algebra.

As a brief interlude, imagine a "calculator" which only has a numeric keypad. What can you do with numbers on such a calculator? Well, you can type them but not much else because the *operations* are missing. So, instead of a calculator we really have a typewriter which, needless to say, is of no use in a mathematical setting. It does not matter how pretty the numbers may look: if we cannot perform arithmetics the numbers do not serve their purpose. Therefore, if we were to define numbers, we would need to define them in terms of operations. But which operations? There are infinitely many operations that one can perform on numbers! For instance, one can raise a number to an integer power, if the number is positive, one can take its square root or logarithm. Clearly, we cannot define numbers by listing all possible operations that can be applied to them: instead, we need to single out the most fundamental *arithmetic* operations.

It is quite clear that addition and multiplication are much more fundamental to arithmetics than, say, natural logarithms. In fact, whatever arithmetic operations one learns in Calculus or elsewhere can always be defined solely in terms of addition and multiplication, e.g.:

$$\ln x = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

This suggests that we should think of numbers as objects that can be added and multiplied: all other number operations are secondary. Furthermore, even if the objects look unnumber-like, it does not matter as long as we have addition and multiplication conforming to the "usual" arithmetic rules: $a+b = b+a$ (commutativity), $c\,(a+b) = c\,a + c\,b$ (distributive law), $a+0 = a$ (additive identity), $1\,a = a$ (multiplicative identity), etc etc.

Let us apply the same philosophy to vectors. The main vector operations are vector addition and scalar multiplication. Hence,

Vectors are objects which can be added and scaled.

The dot product is a very important vector operation which, unlike the cross-product, will prove to be very useful. It is not part of the definition, however, because one can do linear algebra without any reference to the dot product.

Our definition of vectors is not completely rigorous at this point because we do not formally define the operations of vector addition and scaling. To do that, we could follow the Calculus or Physics route and insist on *representing* vectors by arrows or, equivalently, n-tuples of numbers. Then we could formalize vector operations with pictures such as Figure 1 or Equations such as (1) and (2). Advanced Linear Algebra courses start with strict axiomatic definitions which are later used to prove theorems. In a "practical" course, however, it is best to start with an intuitive definition of vectors combined with the intuitive understanding that vector addition and scaling have the "usual" properties, such as: $\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$ (commutativity), $s(\mathbf{a}+\mathbf{b}) = s\,\mathbf{a}+s\,\mathbf{s}$ (distributive law), $\mathbf{a}+\mathbf{0} = \mathbf{a}$ (the existence of the zero vector), etc etc. The important part is that we agree to treat as vector any objects that can be added and scaled as vectors!

Vector spaces

Let us look at several concrete examples of vectors starting with the familiar "geometric" vectors in space. Given our earlier definition of matrix-vector multiplication, we *represent* 3-vectors as columns of numbers:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

On occasion, we will write 3-vectors as linear combinations of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$. However, more often than not, we will simply designate vectors with bold symbols such as $\mathbf{x}$ or $\mathbf{y}$. The set of all 3-vectors is an example of a three-dimensional vector space $\mathbb{R}^3$. As you may have guessed (or already know), the set of all vectors in the plane is labeled $\mathbb{R}^2$ and the set of all n-vectors is labeled $\mathbb{R}^n$.

As our next example, let us consider the set of all quadratics with real coefficients:

$$P_2 = \left\{a + b\,x + c\,x^2 \mid a, b, c \in \mathbb{R}\right\}$$

Students often have a very strong association between quadratics and parabolas which prompts objections to calling quadratics vectors. However, since

$$a_1 + b_1\,x + c_1\,x^2 + a_2 + b_2\,x + c_2\,x^2 = (a_1 + a_2) + (b_1 + b_2)\,x + (c_1 + c_2)\,x^2$$

and

$$s\,(a + b\,x + c\,x^2) = s\,a + s\,b\,x + s\,c\,x^2,$$

the quadratic polynomials add and scale just like vectors and therefore *are* vectors. In fact, we can identify (as we have been doing)

$$a + b\,x + c\,x^2 \equiv \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

which shows that P_2 is algebraically the same, or *isomorphic* to $\mathbb{R}^3$. Likewise, P_n—the set of all polynomials of degree n—is isomorphic to $\mathbb{R}^{n+1}$. In particular, constants P_0 are isomorphic to $\mathbb{R}^1 = \mathbb{R}$: thus we can regard ordinary numbers as single-dimensional vectors.

As a final example, let us consider the set of all two-by-two matrices:

$$M_2 = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a, b, c, d \in \mathbb{R} \right\}$$

Are matrices vectors? Certainly, since they add

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{bmatrix}$$

and scale

$$s \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} s\,a & s\,b \\ s\,c & s\,d \end{bmatrix}$$

just as vectors do. Since a two-by-two matrix is specified by four numbers, the vector space M_2 is four-dimensional. In fact, M_2 is isomorphic to $\mathbb{R}^4$.

Vector spaces and linear transformations

A vector space V is a collection of vectors that is closed under vector operations. That means that if we take any two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V the sum $\mathbf{v}_1 + \mathbf{v}_2$ is also a vector in V; similarly, if we scale some vector $\mathbf{v}$ in V by scalar s, the scaled vector $s\,\mathbf{v}$ will be in V. The earlier examples of $\mathbb{R}^n$, P_n, and M_n are examples of vector spaces. Here we give a few counter-examples.

First, consider $\mathbb{R}^3$ from which the zero vector $\mathbf{0}$ has been removed. The result is a set of vectors S_1 which is not a vector space. Indeed, take any nonzero vector $\mathbf{v} \in S_1$. Then

$$\mathbf{v} + (-1)\,\mathbf{v} = \mathbf{0} \notin S_1$$

It is easy to see that deletion of *any* one vector from $\mathbb{R}^n$ ruins the vector space structure. As another example, let us consider the set S_2 of invertible two-by-two matrices. To see that S_2 is not a vector space, consider

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \notin S_2$$

The identity matrix and its negative are (self)-invertible and are elements of S_2. The zero matrix, however, is clearly not invertible as it sends every vector into the same zero-vector.

A vector space to a linear transformation is what a stage is to an actor[1]: without a stage there can be no performance. Indeed, we say that T is a linear transformation if

$$T(a\,\mathbf{x} + b\,\mathbf{y}) = a\,T(\mathbf{x}) + b\,T(\mathbf{y}).$$

Yet this definition makes sense only if we are allowed to perform vector operations on T's inputs! This explains why the statement $T : X \mapsto Y$ in the context of linear algebra means that T is a linear transformation and X—the domain of T—is a vector space; Y—the target space of T—is also a vector space. As we explain below, Y is not called "range" of T unless certain additional conditions are fulfilled.

[1] Recall our use of expressions like "T acts on" and "the action of T is" to describe linear transformations.

The importance of terminology

Teaching terminology is the most difficult pedagogical aspect of linear algebra. It usually takes a lot of energy to convince the class that the notion of, say, a vector space is both fundamental and really useful. The reason is that the use of a concept is more subtle and very different from, say, the use of the quadratic formula, some trig identity, or any other fact or recipe.

For now, our practical reason for introducing the concept of a vector space is understanding statements such as

$$T : \mathbb{R}^3 \mapsto \mathbb{R}^2$$

or

$$\frac{d}{dx} : P_3 \mapsto P_2,$$

as well as seeing that they may actually be saying the same thing differently. However, very soon we will learn that vector spaces have rich general theory which leads to deep results concerning linear transformations. All of the computational recipes and facts stem from that theory—it is the foundation of Applied Linear Algebra. On this note, here is a list of glossary terms which will become increasingly important as we immerse ourselves into the theory of vector spaces. For completeness, some of the earlier definitions are repeated.

Vectors Objects that add and scale like Calculus and Physics vectors.

Linear combination The operation of scaling and adding a bunch of vectors.

Vector space A set of vectors closed under vector operations: any linear combination of vectors from a vector space is another vector in the same space.

Real vector space A vector space where vectors are scaled by real numbers.

Complex vector space A vector space where scalars are complex numbers.

Vector subspace Any portion of a vector space which is itself a vector space.

Linear transformation A transformation that maps linear combinations of inputs into linear combinations of outputs.

Matrices Rectangular arrays of numbers which can be added and scaled like vectors and, in addition, multiplied "row by column".

Vector representation Identification of objects with vectors in $\mathbb{R}^n$.

Matrix representation Identification of linear transformations with m-by-n matrices.

Domain of a transformation The vector space of its inputs $\mathrm{dom}(T)$.

Co-domain The target space of the operator, that is, a vector space containing all possible outputs.

Injective A [linear] transformation $T : X \mapsto Y$ is called injective, or one-to-one, if $T(x_1) = T(x_2)$ implies that $x_1 = x_2$ for any two vectors in the domain X.

Surjective A [linear] transformation $T : X \mapsto Y$ is surjective, or onto, if for every $y \in Y$ there is at least one $x \in X$ such that $T(x) = y$.

Range The vector space of outputs: not the same as co-domain which is usually bigger than the range.

Operator A transformation $T : X \mapsto X$ whose domain coincides or can be identified with the co-domain.

Isomorphic Literally "built in the same way". There is a formal definition, however, at this point simply think of the conceptual difference between P_2 and $\mathbb{R}^3$—there is not! The only difference between these vector spaces is how we write the vectors: the operations are exactly the same.

Identity Matrix A square matrix with ones on the diagonal and zero off-diagonal elements.

Identity Any transformation which can be represented with the identity matrix.

Inverse A transformation S is the inverse of transformation T if the composition $S \circ T$ is the identity; the same necessarily holds for the composition $T \circ S$.

Invertible A transformation that has an inverse.

Diagonal matrix Any matrix whose off-diagonal elements are zero.

Diagonalization Representation of a linear transformation with a diagonal matrix (one of our main goals!)

Upper (lower) triangular matrix A matrix with zeros below (above) the main diagonal—the next best form after diagonal.

Similar Two matrices are similar if they represent the same linear transformation.

Kernel The kernel of $T : X \mapsto Y$ is the set of all vectors in X which are mapped into the zero vector in Y:

$$\ker(T) = \{\mathbf{x} \in X \mid T(x) = \mathbf{0}\}$$

Exercises

The purpose of the following exercises is twofold: developing familiarity with the linear algebra language and preparing for the general theory that lies ahead. Please read the handout *in full* before attempting.

1. Consider the following sets:

   (a) The set containing the zero-vector $S_1 = \{\mathbf{0}\}$.

   (b) Quadratics without the constant term $S_2 = \{a\,x + b\,x^2 \mid a, b \in \mathbb{R}\}$

   (c) The set of all rational numbers $S_3 = \left\{\frac{m}{n} \mid m, n \in \mathbb{Z}\right\}$ ($\mathbb{Z}$ denotes integers).

   (d) The set of all differentiable functions on the interval $[0, 1]$ satisfying the boundary conditions: $f(0) = f(1) = 0$.

   (e) The set of all differentiable functions on the interval $[0, 1]$ satisfying the boundary conditions: $f(0) = 1, f(1) = 0$.

   Explain clearly in each case why the set is or is not a real vector space.

2. Give an example of four different subspaces of $\mathbb{R}^2$. Explain verbally what makes your examples subspaces.

3. Let T be a linear transformation on some vector space X. Recall that the range of T is the set of all outputs. Prove or disprove: the range of T is a vector space. **Note:** to disprove, provide a counterexample.

4. Let U and V be two vector subspaces of R^2. The intersection $U \cap V$ is the set of all vectors in $\mathbb{R}^2$ which lie both in U and V. Prove or disprove: $U \cap V$ is a subspace of $\mathbb{R}^2$.

5. Suppose $T : X \mapsto Y$ is a transformation with the property $T(x) \neq 0$ for any $x \in X$. Prove or disprove: T is nonlinear.

6. Suppose that $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ has a nontrivial kernel meaning that there exists at least one nonzero vector $x \in \mathbb{R}^2$ such that $T(\mathbf{x}) = \mathbf{0}$. Prove or disprove: T cannot be inverted.