

Chapter 1: Introduction to Vectors

1.1. Vectors and linear combinations

Let's begin by saying what vectors are: They are lists of numbers. If there are 2 numbers in the list, there is a natural correspondence to a point in a plane, determined by the choice of axes. If there are 3 numbers in the list, it corresponds to a point in 3-dimensional space. But there is no reason to stop there — vectors can have 4, or 10, or 2000 entries. Because physics also uses the word “vector”, where they are something like “quantities with magnitude and direction”, it is tempting to think of vectors as arrows. But for us, the only arrows are the ones starting from the origin and ending at the points, used to help us see where the points lie in the plane or in space. We'll make copies of such arrows and move them around in the plane to illustrate various things, but they are artificial copies, not real vectors.

One more technicality: It will usually be convenient for our vectors to be written vertically:

$$\mathbf{v} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$$

There will be times when we use “row vectors” $\mathbf{w} = [3 \ 1 \ 7]$, but most of the time we will use the vertical ones. However, it is harder to type the vertical ones into a book, and they take up more space that way, so our text seems to follow the convention I've seen in other texts: The usual horizontal expression with parens and commas means the vertical list:

$$\mathbf{v} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} = (5, 3, -2) ,$$

and the row vector $\mathbf{u} = [5 \ 3 \ -2]$ is a different object.

So what are the obvious things we can do with vectors? Well, one natural thing, if they have the same number of entries, is to add them, “componentwise” (and we won't try to add them if they don't have the same number of entries):

$$\begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 5+0 \\ 3+1 \\ -2+(-2) \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ -4 \end{bmatrix} .$$

Not very surprising, and it's pretty obvious that vector addition is commutative and associative, as they used to say back in elementary school:

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v} , \quad (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) .$$

Simple enough, but is it really worth looking at? Well, yes, because that is usually what we want to do with vectors: If Aaron needs 2 cups of flour and 1.5 tablespoons of sugar to make a loaf of bread, and Becky needs 1.5 cups of flour and 1 tablespoons of sugar, they will need 3.5 cups of flour and 2.5 tablespoons of sugar to make two loaves of bread

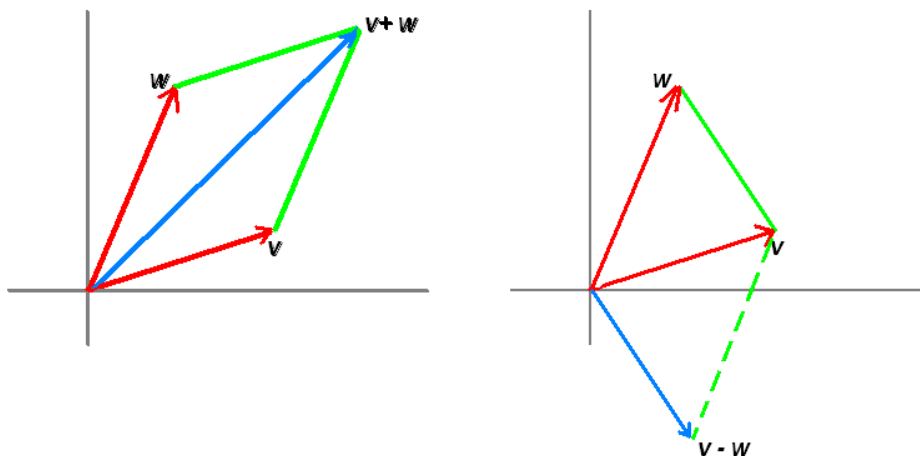
$$\mathbf{a} = \begin{bmatrix} 2 \\ 1.5 \end{bmatrix} , \quad \mathbf{b} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix} : \quad \mathbf{a} + \mathbf{b} = \begin{bmatrix} 3.5 \\ 2.5 \end{bmatrix}$$

— so adding componentwise makes some physical sense.

Subtraction of vectors follows usual rule: $\mathbf{v} - \mathbf{w}$ is the answer to the question “What vector added to $\mathbf{w}$ gives $\mathbf{v}$?” Or in algebra:

$$\mathbf{v} - \mathbf{w} = \mathbf{u} \quad \text{means} \quad \mathbf{w} + \mathbf{u} = \mathbf{v} .$$

When the lists are 2 or 3 components long, and we can represent them geometrically, we often do it by using copies of the vectors at some convenient place:



But again, those copies are not real vectors: The green copies above are not vectors, only the red “operands” (to use the technical term that I may never use again) and the blue results.

By the same token, “multiplication by scalars” makes sense — a “scalar”, in this context, is just a number, and we multiply each component by that scalar, changing the scale of the arrow from the origin to the point:

$$3 \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \\ -6 \end{bmatrix} .$$

Again, it’s what we would want to do with vectors as lists: To make 4 loaves of bread, Aaron needs 8 cups of flour and 6 tablespoons of sugar, and

$$4\mathbf{a} = 4 \begin{bmatrix} 2 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix} .$$

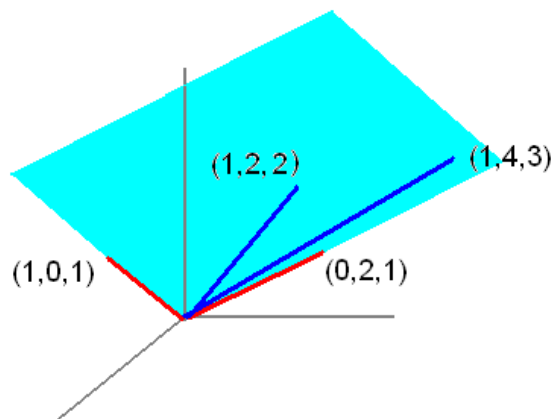
Combining these two operations, we form “(linear) combinations” of vectors:

$$c\mathbf{u} + d\mathbf{v} + e\mathbf{w}$$

where $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are vectors and c, d, e are scalars. It is very important to understand the set of all linear combinations of some vectors. (The text doesn’t use this term until later, but I know I will soon slip and use it, so here goes: The set of all linear combinations of some vectors is called the “*span*” of those vectors.) For a single vector, the set of all combinations is just the set of all scalar multiples of that vector, which form a line through the origin and the vector we started with. For two vectors, if they lie on the same line through the origin, then that line is the set of all their combinations. Moreover, every vector in that line can be written in many different ways as a combination of those vectors:

$$\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} , \quad \mathbf{w} = \begin{bmatrix} 2 \\ -4 \\ -2 \end{bmatrix} : \quad \begin{bmatrix} 3 \\ -6 \\ -3 \end{bmatrix} = \mathbf{v} + \mathbf{w} = 3\mathbf{v} + 0\mathbf{w} = -1\mathbf{v} + 2\mathbf{w} = \dots$$

But if they are not on the same line through the origin, then their combinations form a plane, and every vector in that plane can only be written one way as a combination of the given vectors:



$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} : \quad \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = 1\mathbf{v} + 1\mathbf{w}, \quad \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} = 1\mathbf{v} + 2\mathbf{w}$$

Of course, drawings like this are misleading: The set of all linear combinations is a whole plane through the origin, not a parallelogram — there are no edges — and it is hard to see from the picture that $(1, 2, 2)$ and $(1, 4, 3)$ are in the plane, as is $(-1, 2, 0)$, while $(1, 4, 4)$ is not.

The process continues: If $\mathbf{u}$ is in the plane of combinations of $\mathbf{v}, \mathbf{w}$, then the set of all combinations of all three vectors (in space) is that same plane (and each combination of all three vectors can be written in infinitely many ways). For instance, the set of all combinations of $\mathbf{v} = (1, 0, 1)$, $\mathbf{w} = (0, 2, 1)$ and $\mathbf{u} = (-1, 2, 0)$ is the same plane as for just $\mathbf{v}, \mathbf{w}$, but there are many combinations for a given vector:

$$\begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} = 1\mathbf{v} + 2\mathbf{w} + 0\mathbf{u} = 0\mathbf{v} + 3\mathbf{w} - 1\mathbf{u} = 2\mathbf{v} + 0\mathbf{w} + 1\mathbf{u} = \dots$$

But if $\mathbf{u}$ is not in that plane, then every vector in space is a combination of $\mathbf{v}, \mathbf{w}, \mathbf{u}$, each in only one way. For vectors with 4 entries, the geometry is beyond me, but the algebra goes on as usual: Each one of the vectors, if it is not a combination of the ones before, adds a new dimension to the set of all combinations, and four would “span” all of 4-dimensional space, with every 4-component list written in only one way as a combination of those original four vectors.

It can be hard to decide whether a new vector is a combination of some you already have — that question will be a big chunk of the course shortly. But at least it’s easy to give examples of a set of which all the vectors in $\mathbb{R}^n$ are combinations: The “standard basis” for $\mathbb{R}^n$ is

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0, \dots, 0) \\ \mathbf{e}_2 &= (0, 1, 0, \dots, 0) \\ \mathbf{e}_3 &= (0, 0, 1, \dots, 0) \\ &\vdots \\ \mathbf{e}_n &= (0, 0, 0, \dots, 1) \end{aligned}$$

1.2. Lengths and Dot Products

It may seem natural also to multiply two vectors of the same length componentwise, but it's hard to make sense of that: In the bread-baking example earlier, what would the “product” $(2 \cdot 1.5, 1 \cdot 1) = (3, 1)$ represent? On the other hand, there are good reasons to take that product vector, add its components and get a “dot product” of vectors in which the answer is not a vector but a scalar.

$$(v_1, v_2, \dots, v_n) \cdot (w_1, w_2, \dots, w_n) = v_1 w_1 + v_2 w_2 + \dots + v_n w_n .$$

Notice that the dot product is clearly commutative, but that associativity doesn't make sense, because the dot product of the first two, on either end, isn't a vector anymore. But for a scalar c and vectors $\mathbf{v}$ and $\mathbf{w}$, we do have the rule

$$c(\mathbf{v} \cdot \mathbf{w}) = (c\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (c\mathbf{w}) .$$

WARNING: If and when we try include complex numbers as scalars, we will have to modify both the commutative law and this last rule about scalar multiplication.)

To see why this operation might be good for something, reinterpret the second vector in the bakery example: If a cup of flour costs 1.5 narfs and a tablespoon of sugar 1 narf, then one of Aaron's loaves costs $2(1.5) + 1(1) = 4$ narfs. Another reason this operation might be useful comes from geometry: When the number of entries n is 2 or 3, so that geometry makes sense, then by Pythagoras' theorem, the distance $\|\mathbf{v}\|$ from the origin to $\mathbf{v} = (v_1, \dots, v_n)$, i.e., the “length” of the vector $\mathbf{v}$, is

$$\|\mathbf{v}\| = \sqrt{(v_1 - 0)^2 + \dots + (v_n - 0)^2} = \sqrt{\mathbf{v} \cdot \mathbf{v}} .$$

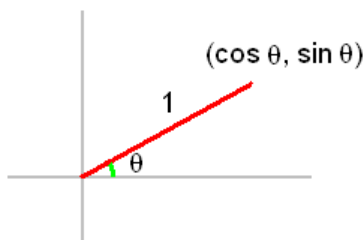
So we can use the dot product for a vector with any number of entries and get a meaning for distance from the origin and/or length of a vector in any “dimension”.

Notice that, for any nonzero vector $\mathbf{v} = (v_1, \dots, v_n)$,

$$\left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| = \left\| \left(\frac{v_1}{\|\mathbf{v}\|}, \dots, \frac{v_n}{\|\mathbf{v}\|} \right) \right\| = \sqrt{\frac{v_1^2}{\|\mathbf{v}\|^2} + \dots + \frac{v_n^2}{\|\mathbf{v}\|^2}} = \sqrt{\frac{v_1^2 + \dots + v_n^2}{\|\mathbf{v}\|^2}} = 1 ,$$

so $\mathbf{v}/\|\mathbf{v}\|$ is a vector of length 1 – a “unit vector” in the direction of $\mathbf{v}$. So, for example, if $\mathbf{v} = (1, 2, -1)$, then $\|\mathbf{v}\| = \sqrt{1^2 + 2^2 + (-1)^2} = \sqrt{6}$, so the unit vector in the same direction as $\mathbf{v}$ is $(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{-1}{\sqrt{6}})$.

In two dimensions, because the coordinates u_1 and u_2 in a unit vector must satisfy $u_1^2 + u_2^2 = 1$, they must be the cosine and sine, respectively of some angle. In fact, that angle is the one that the vector makes with the positive x -axis (or x_1 -axis, or whatever variable is used for the first coordinate):



Here's an algebraic proof of an important, mostly geometric, fact:

$$\begin{aligned} \|\mathbf{v}\|^2\|\mathbf{w}\|^2 - (\mathbf{v} \cdot \mathbf{w})^2 &= (v_1^2 + \cdots + v_n^2)(w_1^2 + \cdots + w_n^2) - (v_1w_1 + \cdots + v_nw_n)^2 \\ &= \sum_i v_i^2w_i^2 + \sum_{i>j} (v_i^2w_j^2 + v_j^2w_i^2) - \sum_i v_i^2w_i^2 - 2 \sum_{i>j} v_iw_iv_jw_j \\ &= \sum_{i>j} (v_iw_j - v_jw_i)^2 \geq 0 \end{aligned}$$

so that we always have

$$|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \|\mathbf{w}\| .$$

This is the Schwarz inequality, or the Cauchy-Buniakowsky-Schwarz inequality, or the CBS inequality, which Strang, with his typical ardor, calls the “most important inequality in mathematics.” My candidate for that honor is the Triangle Inequality, which, I admit, does follow immediately: The distance between two points is at most the sum of the distances from one of them to a third point and then to the second; or, in vector terms:

$$\|\mathbf{v} - \mathbf{w}\| \leq \|\mathbf{v} - \mathbf{u}\| + \|\mathbf{u} - \mathbf{w}\| .$$

(To prove this, first simplify by setting $\mathbf{x} = \mathbf{v} - \mathbf{u}$ and $\mathbf{y} = \mathbf{u} - \mathbf{w}$, so that $\mathbf{v} - \mathbf{w} = \mathbf{x} + \mathbf{y}$, and we want to prove that $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. Then

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = \mathbf{x} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} \\ &= \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2(\|\mathbf{x}\| \|\mathbf{y}\|) + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 . \end{aligned}$$

In fact, if θ denotes the angle between $\mathbf{v}$ and $\mathbf{w}$ at the origin, then

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta .$$

But we won't try to prove this now; we will just note that, if the dot product of two vectors is positive, they form an acute angle; if it is negative, the angle is obtuse.

Question: If $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\|$, what can we say about $\mathbf{v}$ and $\mathbf{w}$? How about if $\mathbf{v} \cdot \mathbf{w} = -\|\mathbf{v}\| \|\mathbf{w}\|$?

And because $\cos 90^\circ = 0$, vectors are perpendicular (at the origin) if and only if their dot product is 0 — we call such vectors “orthogonal”: Because $(1, 2, 3) \cdot (1, 1, -1) = 0$, these are orthogonal, i.e., perpendicular, vectors.

1.3. Matrices

A matrix is just a (2-dimensional) rectangular table of numbers. By tradition, a matrix is denoted by a capital letter, the rows in the matrix are horizontal, the columns are vertical, and the “ (i, j) -th” entry in the matrix A , i.e., the entry in the i -th row and j -th column, is denoted a_{ij} — I actually prefer to include a comma, as in $a_{i,j}$, to avoid confusion in big matrices. WARNING: In computer algebra systems, which are usually case-sensitive, the “ (i, j) -th” entry of A is usually denoted $A(i, j)$ or $A[i, j]$ — you can't switch to lower case.

We are interested in matrices, at least initially, because we can multiply them by vectors and get new vectors, provided the number of columns in the matrix is equal to the number of entries in the vector we multiply by — the product vector will have the number of entries the number of

rows in the matrix. The i -th entry in the product is the result of taking the dot product of the i -th row of the matrix with the given vector. So the process goes like this:

$$A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} : \quad A\mathbf{v} = \begin{bmatrix} 1(3) + 2(4) + 0(5) \\ -1(3) + 1(4) + 2(5) \end{bmatrix} = \begin{bmatrix} 11 \\ 11 \end{bmatrix}.$$

Notice that this product amounts to either

- the column of “dot products” of the rows of A with $\mathbf{v}$, or
- the linear combination of the columns of A with coefficients the entries in $\mathbf{v}$.

$$\begin{bmatrix} (1, 2, 0) \cdot (3, 4, 5) \\ (-1, 1, 2) \cdot (3, 4, 5) \end{bmatrix} \quad \text{or} \quad 3 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 4 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

We can always do the arithmetic to find products like this. The interesting question arises when we know the matrix A and the product vector $\mathbf{b}$, and we want to find the vector $\mathbf{x}$ for which $A\mathbf{x} = \mathbf{b}$ — or even to know whether such a $\mathbf{x}$ exists. If we could “divide by A ”, it would be an easy problem, just like 8th grade algebra; but that’s not always possible, and sometimes not easy even when it is possible. But sometimes it is easy, like when the matrix is “upper triangular” or “lower triangular”:

Example. *Upper triangular matrices:* Take an upper triangular matrix A , i.e., one in which the entries below the “main diagonal”, upper left to lower right, are all zero; and assume that the entries on the main diagonal are nonzero:

$$A = \begin{bmatrix} 3 & -1 & 1 \\ 0 & -1 & -3 \\ 0 & 0 & 2 \end{bmatrix}.$$

Then no matter what 3-entry vector $\mathbf{b}$ is, there will always be exactly one solution: Writing $\mathbf{b} = (r, s, t)$ and $\mathbf{x} = (x, y, z)$ (so that I don’t have to write so many subscripts), we see that $A\mathbf{x} = \mathbf{b}$ amounts to three equations:

$$3x - y + z = r \quad -y - z = s \quad 2z = t$$

We can read off from the third equation that $z = t/2$; substitute into the second equation and find $y = -(s + t/2) = -s - t/2$, and then put that into the first equation to find $x = (r + (-s - t/2) - t/2)/3 = r/3 - s/3 - t/3$. So no matter what $\mathbf{b} = (r, s, t)$ is, we always have only one solution:

$$\mathbf{x} = \begin{bmatrix} r/2 - s/3 - t/3 \\ -s - t/2 \\ t/2 \end{bmatrix}.$$

Thus, *every vector in $\mathbb{R}^3$ is a linear combination of the columns of A* , with exactly one choice for the coefficients.

Notice that this $\mathbf{x}$ is a new matrix times $\mathbf{b}$:

$$\begin{bmatrix} r/2 - s/3 - t/3 \\ -s - t/2 \\ t/2 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/3 & -1/3 \\ 0 & -1 & -1/2 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} r \\ s \\ t \end{bmatrix}.$$

If we call this new matrix A^{-1} , then we have the simple-looking fact that the solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = A^{-1}\mathbf{b}$. We’ll say more about the -1 exponent later, but for now A^{-1} is just another matrix.

If one of the main diagonal entries in this example had been 0, there would not be exactly one solution — there would have been either infinitely many solutions or no solutions at all, depending on what $\mathbf{b}$ was:

Example. If

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad \widehat{\mathbf{b}} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix},$$

then $A\mathbf{x} = \mathbf{b}$ would amount to the three equations

$$x + 2y + z = 1 \quad 2z = 2 \quad z = 2,$$

and it is clear from the last two equations that no choice of z will work in both. But $A\mathbf{x} = \widehat{\mathbf{b}}$ amounts to

$$x + 2y + z = 1 \quad 2z = 2 \quad z = 1,$$

so the last two equations are redundant but solvable, and there are infinitely many choices for x, y that make the first, and hence all three, equations work.

Our results show that $\widehat{\mathbf{b}}$ is a combination of the columns of A — in infinitely many ways — but $\mathbf{b}$ is not a combination of them.

At this point, refer to the presentation for Unit 1.

Chapter 2: Solving Linear Equations

2.1. Vectors and linear equations

2.2. The idea of elimination