COMS10003 : Linear Algebra

Vectors and the Dot Product

Andrew Calway

March 9, 2015

Introduction

This part of the unit is about **linear algebra**. It is called linear algebra because it deals with linear combinations of things. For example, everything that we'll be dealing with can at some level be written in the form of linear equations, such as $a_1x_1 + a_2x_2 + a_3x_3 + .. + a_nx_n = y$. In fact much of linear algebra is about the language that we use to manipulate and reason about these linear forms, without having to explicitly write out all of the details. Thus it makes a lot of use of abstractions and data structures to help us do that.

Why is this important? Well, it turns out that there are many applications in computer science in which linear algebra plays a central role. Examples include search engines, machine learning, algorithms, robotics, graphics, quantum computing and computer vision. Thus computer scientists need to have a good understanding of the key concepts and more importantly, the language of linear algebra.

As was the case with probability, there are many many textbooks and online material about linear algebra. Some are aimed at specific disciplines and applications in science and engineering, including computer science. I suggest you make use of these as much as possible, to get a variety of different explanations and descriptions so that you can make sure that you fully understand the material. As previously, I have made use of several textbooks from my bookshelves and these are listed below.

Theory and problems of linear algebra by Seymour Lipschutz, McGraw-Hill,1981.

Linear Algebra and Probability for CS Applications by Ernest Davis, CRC Press, 2012.

Coding the Matrix by Philip N Klein, Newtonian Press, 2013.

Vectors

We start with one of the basic components of linear algebra - the **vector** - which we have already used when we looked at the gradient and Fourier series. As pointed out then, you'll find that in some textbooks, vectors are treated as abstract entities without concrete definition. In many areas of mathematics this is important. However, we will keep things concrete, as this will make our exploration easier to follow.

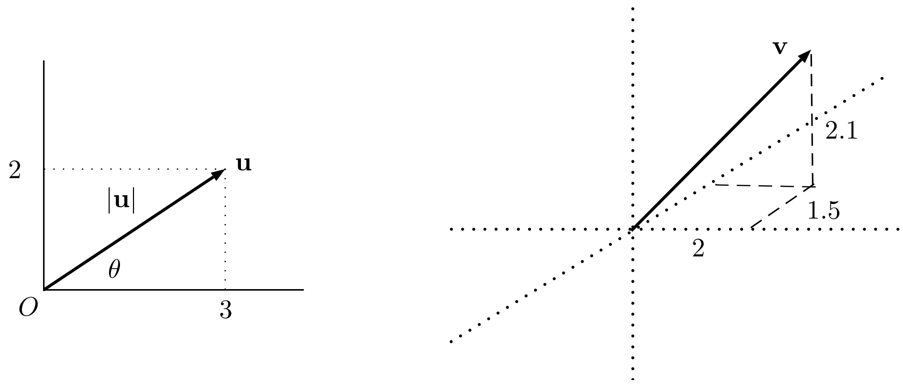

Figure 1: 2-D and 3-D vectors

We will define a vector as an ordered list (or **tuple**) of real numbers. Which order they are in doesn't matter as long as we are consistent and don't change it afterwards. As always, notation varies depending on the discipline and the user - there is no single correct way. We will use round brackets (.) to delimit our ordered list and use bold font to denote a vector. For example, the following are vectors:

$$\mathbf{u} = (3, 2) \qquad \mathbf{v} = (2, -1.5, 2.1) \qquad \mathbf{w} = (0.5, -0.5, -0.5, 0.5)$$

$$\mathbf{x} = (2.1, 3, 4, 5.1, -1.2, 3, 1) \qquad \mathbf{y} = (y_1, y_2, \ldots, y_n)$$

The numbers that make up the vector are known as its **components** and for vector $\mathbf{y}$, say, we denote the ith component by y_i. The above are sometimes known as n-dimensional vectors, where n is the number of components, i.e. $n = 1, 2, 3, ...$, etc. For example, $\mathbf{u}$ and $\mathbf{x}$ are 2-D and 5-D vectors, respectively, and $u_2 = 2$ and $x_6 = 3$.

The set of all such n-dimensional vectors is often denoted by $\mathcal{R}^n$, and is often known as the real coordinate space of n dimensions. Similarly, if the components are complex numbers, then the set of vectors is denoted by $\mathcal{C}^n$.

Magnitude and Direction

In linear algebra, single real numbers are known as **scalars** to distinguish them from vectors. Scalars have a magnitude. Vectors also have a **magnitude** (also called their length) but they also have a **direction**. This is best seen by considering the 2-D vector $\mathbf{u} = (3, 2)$ and representing it on the 2-D plane by an arrow starting at some reference point O and with its endpoint at the coordinates given by the components of the vector as shown on the left in Fig. 1.

In this case the magnitude of the vector can be obtained with help from our Greek friend Mr P, i.e. $|\mathbf{u}| = \sqrt{u_1^2 + u_2^2} = \sqrt{3^2 + 2^2} = \sqrt{13}$, and similarly, the direction can be defined by the angle $\theta = \arctan(u_2/u_1) = \arctan(2/3)$. We can define similar quantities for 3-D vectors and the magnitude for an n-D vector is in general

$$|\mathbf{y}|^2 = y_1^2 + y_2^2 + \ldots + y_n^2 = \sum_{i=1}^{n} y_i^2$$

As another aside, this is also called the **Euclidean norm**. The angle of direction, however, becomes more tricky to define as the number of dimensions increases. We'll return to this later.

We now define two special vectors. The first is the **zero vector** which, unsurprisingly, has zero magnitude and an even less surprising definition, i.e. $\mathbf{O} = (0, 0, 0)$. The second are **unit vectors**, whose magnitudes are equal to one, e.g. for the vector $\mathbf{w} = (0.5, -0.5, -0.5, 0.5)$, $|\mathbf{w}| = \sqrt{0.5^2 + 0.5^2 + 0.5^2 + 0.5^2} = 1$. Note that for any vector $\mathbf{u}$, say, we can define a corresponding unit vector in the same direction as $\mathbf{u}/|\mathbf{u}|$. These vectors are particularly important and we shall use them a lot.

Examples

Having introduced vectors and their basic properties, let us now look at some examples. Vectors are often used to represent physical things, such as force and velocity, and this is probably how you have come across them before in, e.g., physics. In the case of velocity, they can be 2-D or 3-D vectors, with the magnitude representing the speed and the direction representing the direction of travel. As these are physical things they are straightforward to visualise and interpret.

However, as indicated above, we are not restricted to 2-D or 3-D vectors. We can define and work with vectors of any dimension and although these are harder to visualise, they prove to be very useful in many applications in computer science. For example, in document analysis, it is common to make use of what is termed a 'bag of words' model in which documents are represented by vectors in which each component defines the number of occurrences of a given word from a vocabulary in the document.

A simple example might be a vector $(2, 4, 9, \ldots)$ which represents a document with 2 occurrences of the word "sea", 4 of the word "boat", 9 of the word "island", and so on. Thus with a vocabulary of 1000 words, say, documents become vectors in 1000-D space, which is tricky to visualise (don't try, just think of 3-D and relax!), but which proves to be a very powerful representation for applications such as document classification. Take a look at the following web pages: `http://en.wikipedia.org/wiki/Bag-of-words_model` and `http://en.wikipedia.org/wiki/Document_classification`. We shall return to this later.

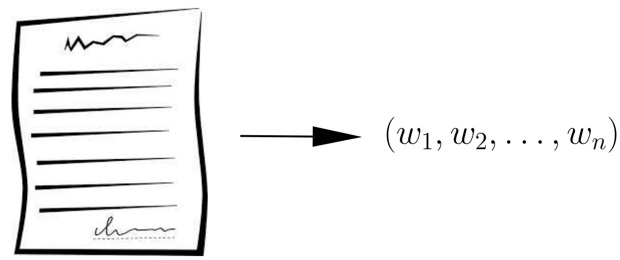

Figure 2: Documents can be represented by vectors, in which the components correspond to the number of occurrences of words from a vocabulary.

Addition and Scalar Multiplication

We now move on to some operations for combining and manipulating vectors. The first one is to add them together and the definition is simple: we add the individual components together to create a new vector, e.g. if $\mathbf{w}_1 = (-1, 2, 3, -5)$ and $\mathbf{w}_2 = (-2, 1, -1, 3)$, then

$$\mathbf{w}_1 + \mathbf{w}_2 = (-1, 2, 3, -5) + (-2, 1, -1, 3) = (-1-2, 2+1, 3-1, -5+3) = (-3, 3, 2, -2)$$

or more generally $\mathbf{y} + \mathbf{z} = (y_1 + z_1, y_2 + z_2, \ldots y_n + z_n)$. It follows that we can only add vectors together if they have the same number of components. The other simple operation is to scale a vector by multiplying it with a scalar, e.g. for a scalar a and vector $\mathbf{y}$, $a\mathbf{y} = (ay_1, ay_2, \ldots ay_n)$, so for the vector $\mathbf{v} = (2, -1.5, 2.1)$, $3\mathbf{v} = (3 \times 2, -3 \times 1.5, 3 \times 2.1) = (6, -4.5, 6.3)$.It should be easy to see that vector addition and scalar multiplication have the same properties as scalar addition and multiplication.

It is useful to visualise both operations as shown in Fig. 3 for the case of 2-D vectors. Note that the addition of two vectors corresponds to the diagonal of the parallelogram formed by the vectors and that scalar multiplication by -1 corresponds to a vector in the opposite direction. The subtraction of two vectors is therefore parallel with and of the same magnitude as the other diagonal of the parallelogram, and the **distance** between the two vectors is given by its magnitude, i.e.

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2}$$

Also note the triangular inequality : the magnitude of the sum of two vectors must be less than or equal to the sum of their magnitudes, i.e. $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$.

Again, it is difficult to visualise, but the same applies to all n-D vectors - the addition of two 10-D vectors, for example, corresponds to the diagonal of the parallelogram formed in their common 2-D plane.

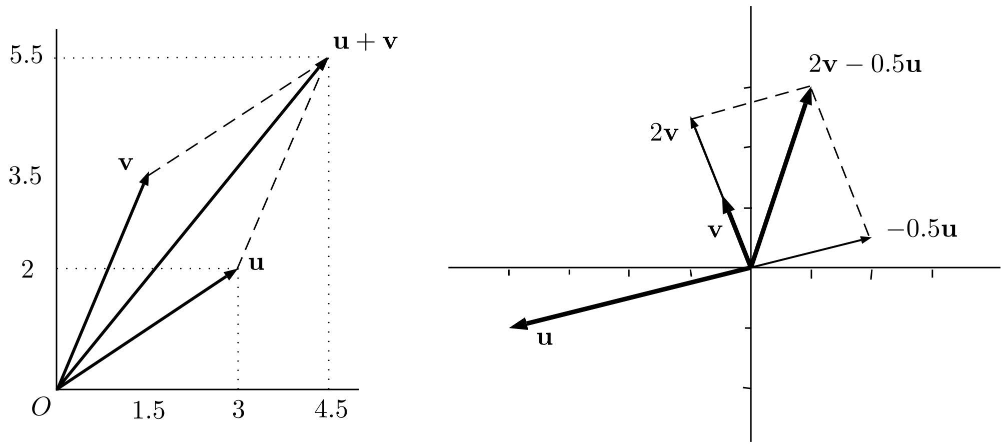

Figure 3: Vector addition and scalar multiplication

The Dot Product

Next we look at another way of combining vectors. It is one way of forming the product of two vectors. Known as the *dot product* - denoted by a '.' - it is defined as follows for two n-D vectors:

$$\mathbf{u}.\mathbf{v} = (u_1v_1 + u_2v_2 + \ldots + u_nv_n)$$

Hence it gives a scalar and corresponds to the sum of the products of the corresponding components -"multiply the components and add em' up". So what does it represent? Let's take a look at the case of two 2-D vectors $\mathbf{u}$ and $\mathbf{v}$. It is straightforward to show (you will prove it in the workshop) that the dot product is also given by

$$\mathbf{u}.\mathbf{v} = |\mathbf{u}|\ |\mathbf{v}|\cos\theta$$

where θ is the angle between the two vectors as shown below. Thus, the dot product indicates how close the vectors are to being parallel - it has a value of 1 or -1 if they are parallel and a value of zero if they are perpendicular or, in the language of linear algebra, **orthogonal**. Note that this applies to all n-D vectors - forming the dot product between two 12-D vectors, for instance, enables us to determine the angle between them, i.e. $\theta = \arccos(\mathbf{u}.\mathbf{v}/|\mathbf{u}||\mathbf{v}|)$.

Note also that if one of the vectors is a unit vector, then the dot product corresponds to the **projection** of the other vector along the direction of the unit vector, i.e. it is the distance along the direction of the unit vector which is closest to the endpoint of the other vector as shown in Fig. 4. In other words, if $\mathbf{u}$ is a unit vector then

$$\mathbf{u}.\mathbf{v} = \arg\min_a d(\mathbf{v}, a\mathbf{u}) \qquad |\mathbf{u}| = 1$$

where $d(,)$ is the distance function. In geometric terms, the vector $a\mathbf{u}$, where $a = \mathbf{u}.\mathbf{v}$, is then orthogonal to the vector $(\mathbf{v} - a\mathbf{u})$ as also shown in Fig. 4.

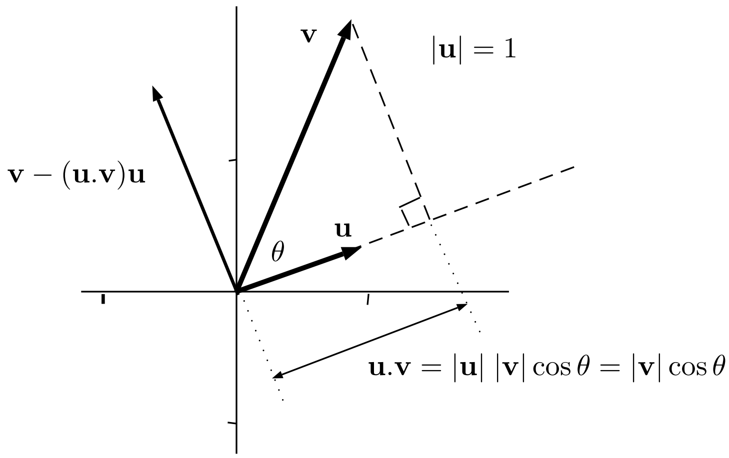

Figure 4: The dot product corresponds to the projection of vector $\mathbf{v}$ onto the direction of unit vector $\mathbf{u}$.

We can also note that the squared magnitude of a vector is given by the dot product of the vector with itself, i.e. $|\mathbf{v}|^2 = \mathbf{v}.\mathbf{v}$, and from the above cosine relationship that $|\mathbf{u}.\mathbf{v}| \leq |\mathbf{u}|\ |\mathbf{v}|$. This is known as the **Cauchy-Schwarz inequality** and makes sense if you consider the geometric interpretation of the dot product. The dot product also has some of the standard properties, e.g. it is commutative and distributive over vector addition, and we'll cover these in the workshop.