Vector Geometry

The main reason to think about vectors (and matrices) is that they are useful in geometry and we will use a lot of geometrical reasoning in computational biology. Here I give a brief introduction to vector geometry. The common treatment is to equate vectors to an ordered set of numbers as we did in the previous handout. But, here I give vectors a bit more abstract treatment that is more directly related to geometry. This will be useful later in the course when we define new notions of distances, etc. It is conventional to use notations like $\vec{x}$ or **x** to distinguish vectors from normal variables that are just numbers (scalars). Here, I will generally use bold face to denote vectors. Finally, formally vector geometry is defined over what is called an algebraic field; but we won't do anything very abstract so all the scalars that we will consider will be just real numbers.

You may remember from high school, vectors as anchored arrows in space. The tips of the arrows are supposed to denote a point in space—thus the connection to geometry. We can add two vectors or subtract two vectors, which then result in another vector; i.e., another point in space.

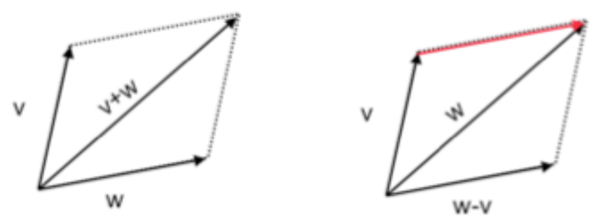

Fig 1.

Fig 1 should remind you of the "parallelogram" rule for vector addition. The picture on the right of Fig 1 shows you that the vector **w** – **v** is parallel to the span of the tips of the vectors **v** and **w** (denoted in red arrow). Thus, if we knew the length of **w** – **v**, we would know the distance separating the tips of **v** and **w**. Standard vector geometry works with points that are defined by the tips of the arrows with the tails all anchored at the same point (the origin). Later on, we might allow the vectors to have their tails anchored at a different point such as the red arrow in the figure above. Allowing for such displacement of the tails gives us what is called an "affine geometry".

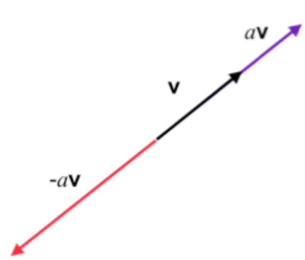

Fig 2

We can also multiply a vector by a scalar (number). In Fig 2, multiplication of vector **v** (black arrow) by a scalar $a > 1$, results in stretching the vector (blue arrow). If $a < 1$, then it would result in a contraction by a certain amount. Furthermore if $a < 0$, then it results in a reflection of the vector with the arrow head in the opposite direction (red arrow). Finally, if $a = 1$, the vector is left alone and if $a = 0$, it results in something called the "**null vector**"—an arrow that has zero length (also called the "**zero vector**").

So, a vector can be added (or subtracted) with another vector to produce a third vector or multiplied with a scalar to produce a new vector. However, addition of a scalar to a vector is not defined.

Geometry is the study of points in space. The important geometrical aspect of vectors is that their multiplication by a scalar and addition of vectors can be used to represent points in space. Suppose we have a non-zero vector, **v**, and consider the one-dimensional line that is an extension of that vector. Other points on that line can be represented by vectors of the form $a\mathbf{v}$, where a is a scalar. We can think of the vector as an arrow with its tail at the origin and all other points along the line are expressed as positive or negative (if the point is in the opposite direction) scalar multiples of **v**.

Fig 3

Suppose now we have two vectors **v** and **w** with $\mathbf{w} \neq a\mathbf{v}$; i.e., **w** is not just an extension of **v**. We can now consider the **2-dimensional space spanned by the two vectors.** Any points within this plane can be written as the sum of the two vectors multiplied by some scalars.

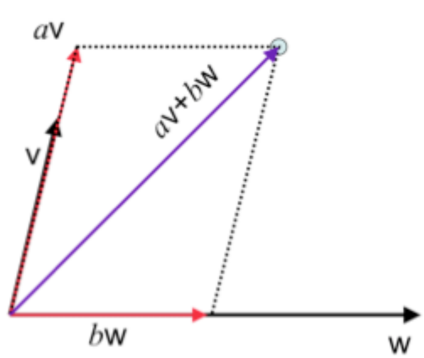

Fig 4

In Fig 4, the point marked by a circle is at the head of the arrow that is the sum of a stretched vector **v** (stretched by a amount) and a contracted vector **w** (contracted by b amount). That is, for any vector (and associated point) **x** within the plane spanned by **v** and **w**, we can write,

$$\mathbf{x} = a\mathbf{v} + b\mathbf{w}$$

for some numbers a and b. This tells us that we can conveniently work with just **v** and **w** plus some scalars for any vector within the plane spanned by **v** and **w**. Conversely, if we choose two non-zero distinct vectors **v** and **w**, the vectors define a set of points that is a two-dimensional plane and we say that the **v** and **w** form the **basis vectors** for the plane.

If we have p distinct vectors, by which we mean none of the vectors can be written as a linear sum of the other vectors, then we can denote all the points within a p-dimensional space as,

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_p\mathbf{v}_p$$

where $\mathbf{v_1} \cdots \mathbf{v_p}$ are the basis vectors. Therefore, a collection of vectors can be used to represent a collection of points and this is what we mean by a **vector space**. (And, also what we mean by "dimensions".)

In geometry, we have two fundamental quantities of interest: lengths and angles. We now show a way to define lengths and angles from vectors. We define something called **inner products** of two vectors and use the following notation:

$$< \mathbf{v}, \mathbf{w} >$$

Inner products of two vectors are functions of two vectors that return a real value—that is a number (scalar). Input two vectors, get one number. You can also think of this as a computational operation on two vectors that results in a scalar. These inner products are defined in such a way to have the following properties:

1. Symmetry: $< \mathbf{v}, \mathbf{w} > = < \mathbf{w}, \mathbf{v} >$
2. Positive-definiteness: $< \mathbf{v}, \mathbf{v} > > 0$
3. Bilinearity: $\begin{cases} < a\mathbf{v}, \mathbf{w} > = a < \mathbf{v}, \mathbf{w} > \\ < \mathbf{v} + \mathbf{x}, \mathbf{w} > = \quad < \mathbf{v}, \mathbf{w} > + < \mathbf{x}, \mathbf{w} > \\ < \mathbf{v}, \mathbf{w} + \mathbf{x} > = \quad < \mathbf{v}, \mathbf{w} > + < \mathbf{v}, \mathbf{x} > \\ < \mathbf{v}, a\mathbf{w} + b\mathbf{x} > = a < \mathbf{v}, \mathbf{w} > + b < \mathbf{v}, \mathbf{x} > \end{cases}$ (*)

(The property definition (*) given above is a bit of a simplification for the case when the scalars are restricted to real numbers, as opposed to things like complex numbers.)

The notation in property #2 is a bit confusing because I used brackets for the inner products, which can be confused with the "greater than" symbol. It just says that if you take the inner product of the same vector, it is always greater than zero. Property #3 shows some examples of what we mean by bilinearity—that is, when we input sums of vectors or scalar multiples of vectors, the operation behaves as if we were distributing the inner product operation with vector additions and (scalar) multiplications.

The inner products can now be used to define the notion of a length of a vector:

$$\|\mathbf{v}\| = \sqrt{< \mathbf{v}, \mathbf{v} >} \qquad \text{(Eq 1)}$$

That is, compute the inner product of the vector with itself (which then returns a number) and take the square root of that. We will also call this quantity (Eq 1), the **norm of the vector** (which is a mathematical way of saying "length").

We now define a **unit vector** as a vector whose norm is 1; i.e., has a unit length of 1. If I have a vector **v**, then its unit vector is:

$\mathbf{v}/\|\mathbf{v}\|$ (Eq 2)

That is, the vector multiplied by a scalar that is the inverse of its norm. We can do a trivial check to see that the new vector does have unit length:

$$\sqrt{\left\langle \frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\rangle} = \sqrt{\frac{\mathbf{1}}{\|\mathbf{v}\|}\left\langle \mathbf{v}, \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\rangle} = \sqrt{\frac{\mathbf{1}}{\|\mathbf{v}\|^2} < \mathbf{v},\mathbf{v} >} = \frac{\sqrt{< \mathbf{v},\mathbf{v} >}}{\|\mathbf{v}\|} = 1 \qquad \text{(Eq 3)}$$

Eq 3 is trivial but shows you how the properties (*) can be used to make sure what we define make sense. (You should make sure you understand the steps of Eq 3 to be comfortable with these kinds of abstract manipulations.). Geometrically, a unit vector of **v** is another vector that has the same direction as **v** and lies in the same line as **v**. (See also Fig 3.)

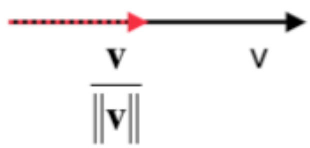

Fig 5

Unit vectors are very useful because of the following computational convenience. Suppose **v** is a unit vector so that $\|\mathbf{v}\| = 1$ and we would like to find the vector in the **v** direction whose length is 4.32. That is, we would like to find the vector $a\mathbf{v}$ such that $\|a\mathbf{v}\| = 4.32$. Then it turns out a = 4.32. (Why?)

In general, if **v** is a unit vector, a vector whose length is m is $m\mathbf{v}$ (Eq 4)

We can now define angles between two vectors, **v** and **w** (Fig 6):

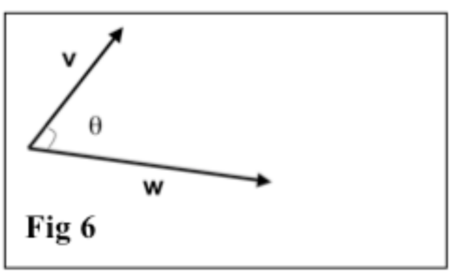

Fig 6

$$\cos\theta = \frac{< \mathbf{v},\mathbf{w} >}{\|\mathbf{v}\|\|\mathbf{w}\|} = \left\langle \frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{\mathbf{w}}{\|\mathbf{w}\|} \right\rangle$$
$$\rightarrow \theta = \cos^{-1}\left(\frac{< \mathbf{v},\mathbf{w} >}{\|\mathbf{v}\|\|\mathbf{w}\|} \right) \qquad \text{(Eq 5)}$$

where $\cos^{-1}$ is the inverse of the cos function called arccos (recall from high school trigs). The second formula in the first line of Eq 5 tells us that the cosine of the angle between two vectors can be obtained by the inner product of the unit vectors in the direction of **v** and **w**. Therefore, the inner product gives us both lengths and angles, all the quantities necessary to quantify geometry.

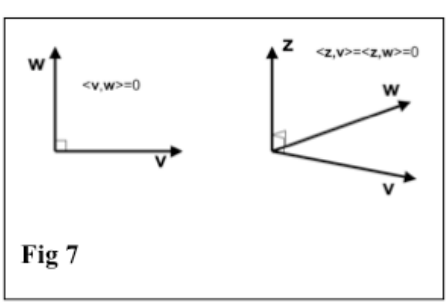

Fig 7

Two non-zero vectors, **v** and **w**, are defined as orthogonal if <**v**,**w**> = 0 (Fig 7). This is immediate from Eq 4 where $\cos\theta = 0 \Rightarrow \theta = \pi/2 = 90^0$. More generally, we can define orthogonal vectors to a p-dimensional space spanned by a set of vectors. Fig 7 shows a vector (in 3-dimensional space) where the vector **z** is orthogonal to **v** and **w**. We use a slightly more general term and say "**z** is **normal** to the space spanned by **v** and **w**", meaning that $<\mathbf{z}, a\mathbf{w}+b\mathbf{v}> = 0$ for all a and b. (Don't confuse a normal vector with the *norm of a vector*.)

The next is a very important property that we will use many times in this course.

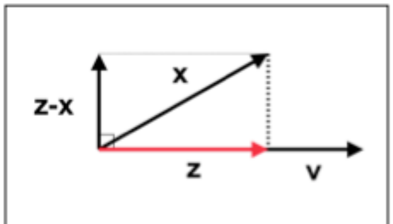

Fig 8

Let **v** and **x** be two non-zero vectors. We would like to obtain another vector **z** that is the **projection** of **x** onto **v.** By a projection, we mean that we would like to extend an orthogonal line from the tip of **x** onto the line along **v** direction (see Fig 8). This means that the vector **z-x** is orthogonal to **z**. See Fig 8.
Since **z** is along the **v** direction, $\mathbf{z} = a\mathbf{v}$ for some unknown scalar a. We can use trig relations to find a. Recall from high school that,

$$\cos\theta = \|\mathbf{z}\|/\|\mathbf{x}\| \qquad \text{(Eq 6)}$$

and from our definition of angles (Eq 5),

$$\cos\theta = \frac{<\mathbf{x},\mathbf{v}>}{\|\mathbf{x}\|\|\mathbf{v}\|} \qquad \text{(Eq 7)}$$

We set Eq 5 equal to Eq 6, do a little manipulation and get,

$$\begin{aligned} &\frac{\|\mathbf{z}\|}{\|\mathbf{x}\|} = \frac{<\mathbf{x},\mathbf{v}>}{\|\mathbf{x}\|\|\mathbf{v}\|} \\ \Rightarrow \quad &\|\mathbf{z}\| = \frac{<\mathbf{x},\mathbf{v}>}{\|\mathbf{v}\|} = \left\langle \mathbf{x}, \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\rangle \end{aligned} \qquad \text{(Eq 8)}$$

Eq 8 tells us that the length of the desired vector **z** can be obtained from taking the unit vector in the **v** direction and computing the inner product with **x**. So,

$$a = \frac{\|\mathbf{z}\|}{\|\mathbf{v}\|} = \frac{\mathbf{1}}{\|\mathbf{v}\|}\left\langle \mathbf{x}, \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\rangle \qquad \text{(Eq 9)}$$

Things look a lot simpler if **v** was a unit vector in the first place. Then $\|\mathbf{v}\| = 1$ and

$$\mathbf{z} = <\mathbf{x},\mathbf{v}>\mathbf{v} \qquad \text{(Eq 10)}$$

Eq 10 is very important for future computations and our geometrical intuition. First, lets be careful in reading the notation. The first part $<\mathbf{x},\mathbf{v}>$ is the inner product of **x** and **v** and the result is a scalar. Then we take this scalar and multiply it to **v**, which is now a new vector in the direction of **v** and we have the notation <**x,v**> **v**.

Recall now the idea of a coordinate system. This idea was invented by R. Descartes (thus the Cartesian coordinate system). Given a point in space, the idea is to represent the point by a set of numbers, i.e., a coordinate system, where each coordinate number is gotten by sending a perpendicular line to reference axes and recording the distance of the intersection point from the origin (cf., Fig 9). Legend goes (probably wrong) that Descartes was lying in bed looking at a fly on the ceiling trying to figure out how to describe its location. Then he hit upon the idea of extending an imaginary line to the edges of the ceiling. Using this idea, every point can be assigned a set of numbers. Suppose that we want two distinct unit vectors **v** and **w** to be the axis of our coordinate system for points in 2-dimensional space as shown in Fig 9 (the unit circle is meant to show that **v** and **w** are unit vectors). Now consider the point represented by the vector **x**. We can immediately see from the discussion leading to Eq 10, that the perpendicular point to **v** is <**x,v**>**v** and the perpendicular point to **w** is <**x,w**>**w** (see Fig 9). That is, the point **x** is represented by the ordered set of numbers (<**x,v**>,<**x,w**>).

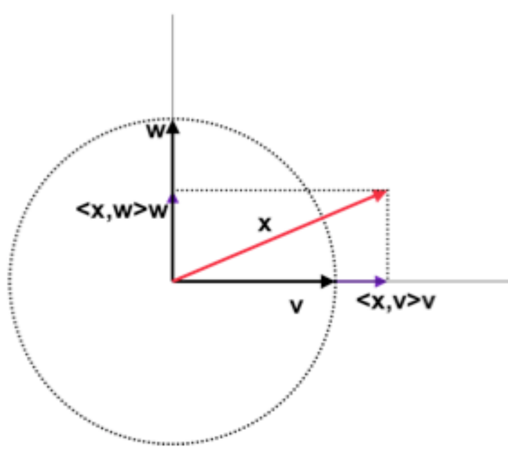

Fig 9

Recall the discussion above about points spanned by *p* vectors. The above discussion suggests that

$$\mathbf{x} = a\mathbf{v} + b\mathbf{w} \qquad \text{(Eq 11)}$$

for some *a* and *b*. Suppose that **v** and **w** are orthogonal as in Fig 9. Then the parallelogram vector addition rule immediately suggests that

$$\begin{aligned} a &= \langle \mathbf{x},\mathbf{v} \rangle \\ b &= \langle \mathbf{x},\mathbf{w} \rangle \end{aligned} \qquad \text{(Eq 12)}$$

So, if **v** and **w** are unit vectors and orthogonal, we find that any vector **x** that is in the plane spanned by **v** and **w** have the coordinates (<**x**,**v**>, <**x**,**w**>) with respect to **v** and **w** and can be written as the vector **x** = <**x**,**v**> **v** + <**x**,**w**> **w**. When a set of *p* vectors is composed of unit vectors that are mutually orthogonal to each other, we will call such a set the **orthonormal basis** for a *p*-dimensional vector space. Everything we have done will be consistent even if the vectors were not unit vectors nor orthogonal to each other. However, in that case formulas such as Eq 10 or Eq 12 have to be modified and the results are messier. Working on vector geometry with respect to an orthonormal basis makes the computations clean as can be seen below.

(At this point, it will be good to pause and make sure that we are confident of what quantities are vectors and what quantities are scalars in all of the discussion above and re-read everything.)

We defined inner products and worked with them but up to now we didn't do any concrete computations—we left the operations as <**v,w**>, etc. When we represent vectors with a coordinate system as discussed just above, then we replace vectors with their numerical representations such as (<x,y>, <x,w>). It turns out then we can also compute the inner products purely in terms of the coordinates. Let **v** and **w** be orthonormal basis of a 2-dimensional space. Then, two vectors **x** and **y** have the following representation:

$$\begin{aligned} \mathbf{x} &= <\mathbf{x},\mathbf{v}>\mathbf{v} + <\mathbf{x},\mathbf{w}>\mathbf{w} \\ \mathbf{y} &= <\mathbf{y},\mathbf{v}>\mathbf{v} + <\mathbf{y},\mathbf{w}>\mathbf{w} \end{aligned} \qquad \text{(Eq 13)}$$

To save some notation, we set a = <**x,v**>, b = <**x,w**>, c = <**y,v**>, d= <**y,w**> and we compute <**x,y**>:

$$\begin{aligned} <\mathbf{x},\mathbf{y}> &= \langle a\mathbf{v}+b\mathbf{w}, c\mathbf{v}+d\mathbf{w}\rangle \\ &= ac<\mathbf{v},\mathbf{v}> + ad<\mathbf{v},\mathbf{w}> + bc<\mathbf{w},\mathbf{v}> + bd<\mathbf{w},\mathbf{w}> \qquad \text{(Eq 14)} \\ &= ac+bd \end{aligned}$$

The second line comes from the bilinear property of (*) and the third line is because $<\mathbf{v},\mathbf{v}>=<\mathbf{w},\mathbf{w}>=1$ from unit vectors and $<\mathbf{v},\mathbf{w}>=<\mathbf{w},\mathbf{v}>=0$ from the fact that **v** and **w** are orthogonal. So, <**x,y**> is obtained by taking the product of each coordinate and summing all the coordinate-wise products. If we remember back to our notation for *a,b,c*, and *d*, then it turns out,

$$<\mathbf{x},\mathbf{y}> \;=\; <\mathbf{x},\mathbf{v}><\mathbf{y},\mathbf{v}> + <\mathbf{x},\mathbf{w}><\mathbf{y},\mathbf{w}> \qquad \text{(Eq 15)}$$

The busy notation of Eq 15 is a reminder to us to keep track of what are vectors and what are scalars. Regardless, the main result we get is that if **v** and **w** are orthonormal vectors then we can compute the inner product between any two vectors in the 2-dimensional space represented by **v** and **w** by simply taking products and sums of the respective coordinates.

Here is now the important fact. Suppose that we are told the vector **x** has coordinates $(x_1,\cdots x_p)$ and vector **y** has coordinates $(y_1,\cdots y_p)$ with respect to some orthonormal set of basis vectors. Then, by the results we have seen in Eq 14 and Eq 15,

$$<\mathbf{x},\mathbf{y}>=\sum_{i=1}^{p} x_i y_i \qquad \text{(Eq 16)}$$

Eq 16 is the "dot product" that some of you may have already learned. It turns out that the dot product is an inner product with respect to orthonormal basis vectors—which we call the Euclidean coordinate system. Once we adopt a coordinate system, all the abstract manipulations we did above can be done concretely with number computations on the coordinate system. A Euclidean coordinate system allows clean formulas but a coordinate system need not be orthonormal. If it isn't then Eq 16 becomes modified by a little bit. (Think about what would happen in Eq 14 if <**v,w**> isn't zero.)

The reason we didn't go straight to Eq 16 is that later in the course we have to deal with cases when we might consider more general kinds of basis vectors. The formalism we covered gives us the fundamentals to understand the more general concepts of space covered in machine learning topics.