Math 115A - Week 9
Textbook sections: 6.1-6.2
Topics covered:

- Orthogonality
- Orthonormal bases
- Gram-Schmidt orthogonalization
- Orthogonal complements

* * * * *

Orthogonality

- From your lower-division vector calculus you know that two vectors v, w in $\mathbf{R}^2$ or $\mathbf{R}^3$ are perpendicular if and only if $v \cdot w = 0$; for instance, $(3, 4)$ and $(-4, 3)$ are perpendicular.
- Now that we have inner products - a generalization of dot products - we can now give a similar notion for all inner product spaces.
- **Definition.** Let V be an inner product space. If v, w are vectors in V, we say that v and w are *orthogonal* if $\langle v, w\rangle = 0$.
- **Example.** In $\mathbf{R}^4$ (with the standard inner product), the vectors $(1, 1, 0, 0)$ and $(0, 0, 1, 1)$ are orthogonal, as are $(1, 1, 1, 1)$ and $(1, -1, 1, -1)$, but the vectors $(1, 1, 0, 0)$ and $(1, 0, 1, 0)$ are not orthogonal. In $\mathbf{C}^2$, the vectors $(1, i)$ and $(1, -i)$ are orthogonal, but $(1, 0)$ and $(i, 0)$ are not.
- **Example.** In any inner product space, the 0 vector is orthogonal to everything (why?). On the other hand, a non-zero vector cannot be orthogonal to itself (why? Recall that $\langle v, v\rangle = \|v\|^2$).
- **Example.** In $C([0, 1]; \mathbf{C})$ with the inner product

$$\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}\ dx,$$

the functions 1 and $x - \frac{1}{2}$ are orthogonal (why?), but 1 and x are not. However, in $C([-1,1];\mathbf{C})$ with the inner product

$$\langle f, g\rangle := \int_{-1}^{1} f(x)\overline{g(x)}\, dx,$$

the functions 1 and $x - \frac{1}{2}$ are no longer orthogonal, however the functions 1 and x now are. Thus the question of whether two vectors are orthogonal depends on which inner product you use.

- Sometimes we say that v and w are *perpendicular* instead of *orthogonal*. This makes the most sense for $\mathbf{R}^n$, but can be a bit confusing when dealing with other inner product spaces such as $C([-1,1],\mathbf{C})$ - how would one visualize the functions 1 and x being "perpendicular", for instance (or i and x, for that matter)? So I prefer to use the word orthogonal when dealing with general inner product spaces.

- Sometimes we write $v \perp w$ to denote the fact that v is orthogonal to w.

- Being orthogonal is at the opposite extreme of being parallel; recall from the Cauchy-Schwarz inequality that $|\langle v, w\rangle|$ must lie between 0 and $\|v\|\|w\|$. When v and w are parallel then $|\langle v, w\rangle|$ attains its maximum possible value of $\|v\|\|w\|$, while when v and w are orthogonal then $|\langle v, w\rangle|$ attains its minimum value of 0.

- Orthogonality is symmetric: if v is orthgonal to w then w is orthogonal to v. (Why? Use the conjugate symmetry property and the fact that the conjugate of 0 is 0).

- Orthogonality is preserved under linear combinations:

- **Lemma 1.** Suppose that $v_1, \ldots, v_n$ are vectors in an inner product space V, and suppose that w is a vector in V which is orthogonal to all of $v_1, v_2, \ldots, v_n$. Then w is also orthogonal to any linear combination of $v_1, \ldots, v_n$.

- **Proof** Let $a_1v_1 + \ldots + a_nv_n$ be a linear combination of $v_1, \ldots, v_n$. Then by linearity

$$\langle a_1v_1 + \ldots + a_nv_n, w\rangle = a_1\langle v_1, w\rangle + \ldots + a_n\langle v_n, w\rangle.$$

But since w is orthogonal to each of $v_1, \ldots, v_n$, all the terms on the right-hand side are zero. Thus w is orthogonal to $a_1 v_1 + \ldots + a_n v_n$ as desired. $\square$

- In particular, if v and w are orthogonal, then cv and w are also orthogonal for any scalar c (why is this a special case of Lemma 1?)

- You are all familiar with the following theorem about orthogonality.

- **Pythagoras's theorem** If v and w are orthogonal vectors, then $\|v + w\|^2 = \|v\|^2 + \|w\|^2$.

- **Proof.** We compute

$$\|v + w\|^2 = \langle v + w, v + w \rangle = \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle.$$

But since v and w are orthogonal, $\langle v, w \rangle$ and $\langle w, v \rangle$ are zero. Since $\langle v, v \rangle = \|v\|^2$ and $\langle w, w \rangle = \|w\|^2$, we obtain $\|v + w\|^2 = \|v\|^2 + \|w\|^2$ as desired. $\square$

- This theorem can be generalized:

- **Generalized Pythagoras's theorem** If $v_1, v_2, \ldots, v_n$ are all orthogonal to each other (i.e. $v_i \perp v_j = 0$ for all $i \neq j$) then

$$\|v_1 + v_2 + \ldots + v_n\|^2 = \|v_1\|^2 + \|v_2\|^2 + \ldots + \|v_n\|^2.$$

- **Proof.** We prove this by induction. If $n = 1$ the claim is trivial, and for $n = 2$ this is just the ordinary Pythagoras theorem. Now suppose that $n > 2$, and the claim has already been proven for $n - 1$. From Lemma 1 we know that v_n is orthogonal to $v_1 + \ldots + v_{n-1}$, so

$$\|v_1 + v_2 + \ldots + v_n\|^2 = \|v_1 + \ldots + v_{n-1}\|^2 + \|v_n\|^2.$$

On the other hand, by the induction hypothesis we know that

$$\|v_2 + \ldots + v_{n-1}\|^2 = \|v_2\|^2 + \ldots + \|v_{n-1}\|^2.$$

Combining the two equations we obtain

$$\|v_1 + v_2 + \ldots + v_n\|^2 = \|v_1\|^2 + \|v_2\|^2 + \ldots + \|v_n\|^2$$

as desired. $\square$

- Recall that if two vectors are orthogonal, then they remain orthogonal even when you multiply one or both of them by a scalar. So we have

- **Corollary 2.** If $v_1, v_2, \ldots, v_n$ are all orthogonal to each other (i.e. $v_i \perp v_j$ for all $i \neq j$) and $a_1, \ldots, a_n$ are scalars, then

  $$\|a_1 v_1 + a_2 v_2 + \ldots + a_n v_n\|^2 = |a_1|^2 \|v_1\|^2 + |a_2|^2 \|v_2\|^2 + \ldots + |a_n|^2 \|v_n\|^2.$$

- **Definition** A collection $(v_1, v_2, \ldots, v_n)$ of vectors is said to be *orthogonal* if every pair of vectors is orthogonal to each other (i.e. $\langle v_i, v_j \rangle = 0$ for all $i \neq j$). If a collection is orthogonal, and furthermore each vector has length 1 (i.e. $\|v_i\| = 1$ for all i) then we say that the collection is *orthonormal*.

- **Example** In $\mathbf{R}^4$, the collection $((3,0,0,0), (0,4,0,0), (0,0,5,0))$ is orthogonal but not orthonormal. But the collection $((1,0,0,0), (0,1,0,0), (0,0,1,0))$ is orthonormal (and therefore orthogonal). Note that any single vector v_1 is always considered an orthogonal collection (why?).

- **Corollary 3**. If $(v_1, v_2, \ldots, v_n)$ is an orthonormal collection of vectors, and $a_1, \ldots, a_n$ are scalars, then

  $$\|a_1 v_1 + a_2 v_2 + \ldots + a_n v_n\|^2 = |a_1|^2 + |a_2|^2 + \ldots + |a_n|^2.$$

  Note that the right-hand side $|a_1|^2 + |a_2|^2 + \ldots + |a_n|^2$ is always positive, unless $a_1, \ldots, a_n$ are all zero. Thus $a_1 v_1 + \ldots + a_n v_n$ is always non-zero, unless $a_1, \ldots, a_n$ are all zero. Thus

- **Corollary 4**. Every orthonormal collection of vectors is linearly independent.

* * * * *

Orthonormal bases

- As we have seen, orthonormal collections of vectors have many nice properties. As we shall see, things are even better when this collection is also a basis:

- **Definition** An *orthonormal basis* of an inner product space V is a collection $(v_1, \ldots, v_n)$ of vectors which is orthonormal and is also an ordered basis.

- **Example.** In $\mathbf{R}^4$, the collection $((1,0,0,0),(0,1,0,0),(0,0,1,0))$ is orthonormal but is not a basis. However, the collection $((1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1))$ is an orthonormal basis. More generally, the standard ordered basis of $\mathbf{R}^n$ is always an orthonormal basis, as is the standard ordered basis of $\mathbf{C}^n$. (Actually, the standard bases of $\mathbf{R}^n$ and of $\mathbf{C}^n$ are the same; only the field of scalars is different). The collection $((1,0,0,0),(1,1,0,0),(1,1,1,0),(1,1,1,1))$ is a basis of $\mathbf{R}^4$ but is not an orthonormal basis.

- From Corollary 4 we have

- **Corollary 5** Let $(v_1, \ldots, v_n)$ be an orthonormal collection of vectors in an n-dimensional inner product space. Then $(v_1, \ldots, v_n)$ is an orthonormal basis.

- **Proof.** This is just because any n linearly independent vectors in an n-dimensional space automatically form a basis. $\square$

- **Example** Consider the vectors $(3/5, 4/5)$ and $(-4/5, 3/5)$ in $\mathbf{R}^2$. It is easy to check that they have length 1 and are orthogonal. Since $\mathbf{R}^2$ is two-dimensional, they thus form an orthonormal basis.

- Let $(v_1, \ldots, v_n)$ be an ordered basis of an n-dimensional inner product space V. Since $(v_1, \ldots, v_n)$ is a basis, we know that every vector v in V can be written as a linear combination of $v_1, \ldots, v_n$:

$$v = a_1 v_1 + \ldots + a_n v_n.$$

In general, finding these scalars $a_1, \ldots, a_n$ can be tedious, and often requires lots of Gaussian elimination. (Try writing $(1,0,0,0)$ as a linear combination of $(1,1,1,1)$, $(1,2,3,4)$, $(2,2,1,1)$ and $(1,2,1,2)$, for instance). However, if we know that the basis is an *orthonormal* basis, then finding these coefficients is much easier.

- **Theorem 6.** Let $(v_1, \ldots, v_n)$ be an orthonormal basis of an inner product space V. Then for any vector $v \in V$, we have

$$v = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n$$

where the scalars $a_1, \ldots, a_n$ are given by the formula

$$a_j = \langle v, v_j \rangle \text{ for all } j = 1, \ldots, n$$

- **Proof.** Since $(v_1, \ldots, v_n)$ is a basis, we know that $v = a_1 v_1 + \ldots + a_n v_n$ for *some* scalars $a_1, \ldots, a_n$. To finish the proof we have to solve for $a_1, \ldots, a_n$ and verify that $a_j = \langle v, v_j \rangle$ for all $j = 1, \ldots, n$. To do this we take our equation for v and take inner products of both sides with v_j:
$$\langle v, v_j \rangle = \langle a_1 v_1 + a_2 v_2 + \ldots + a_n v_n, v_j \rangle.$$
We expand out the right-hand side as
$$a_1 \langle v_1, v_j \rangle + a_2 \langle v_2, v_j \rangle + \ldots + a_n \langle v_n, v_j \rangle.$$
Since $v_1, \ldots, v_n$ are orthogonal, all the inner products vanish except $\langle v_j, v_j \rangle = \|v_j\|^2$. But $\|v_j\| = 1$ since $v_1, \ldots, v_n$ is also orthonormal. So we get
$$\langle v, v_j \rangle = 0 + \ldots + 0 + a_j \times 1 + 0 + \ldots + 0$$
as desired. $\square$

  From the definition of co-ordinate vector $[v]^\beta$, we thus have a simple way to compute co-ordinate vectors:

- **Corollary 7** Let $\beta = (v_1, \ldots, v_n)$ be an orthonormal basis of an inner product space V. Then the co-ordinate vector $[v]^\beta$ of any vector v is then given by
$$[v]^\beta = \begin{pmatrix} \langle v, v_1 \rangle \\ \langle v, v_2 \rangle \\ \vdots \\ \langle v, v_n \rangle \end{pmatrix}.$$

- Note that Corollary 7 also gives us a reltaively quick way to compute the co-ordinate matrix $[T]_\beta^\gamma$ of a linear operator $T : V \to W$ provided that γ is an orthonormal basis, since the columns of $[T]_\beta^\gamma$ are just $[Tv_j]^\gamma$, where v_j are the basis vectors of β.

- **Example.** Let $v_1 := (1,0,0)$, $v_2 := (0,1,0)$, $v_3 := (0,0,1)$, and $v := (3,4,5)$. Then (v_1, v_2, v_3) is an orthonormal basis of $\mathbf{R}^3$. Thus

$$v = a_1 v_1 + a_2 v_2 + a_3 v_3,$$

where $a_1 := \langle v, v_1 \rangle = 3$, $a_2 := \langle v, v_2 \rangle = 4$, and $a_3 := \langle v, v_3 \rangle = 5$. (Of course, in this case one could expand v as a linear combination of v_1, v_2, v_3 just by inspection.)

- **Example.** Let $v_1 := (3/5, 4/5)$, $v_2 := (-4/5, 3/5)$, and $v := (1,0)$. Then (v_1, v_2) is an orthonormal basis for $\mathbf{R}^2$. Now suppose we want to write v as a linear combination of v_1 and v_2. We could use Gaussian elimination, but because our basis is orthogonal we can use Theorem 6 instead to write

$$v = a_1 v_1 + a_2 v_2$$

where $a_1 := \langle v, v_1 \rangle = 3/5$ and $a_2 := \langle v, v_2 \rangle := -4/5$. Thus $v = \frac{3}{5}v_1 - \frac{4}{5}v_2$. (Try doing the same thing using Gaussian elimination, and see how much longer it would take!). Equivalently, we have

$$[v]^{(v_1,v_2)} = \begin{pmatrix} 3/5 \\ -4/5 \end{pmatrix}.$$

- **The example of Fourier series.** We now give an example which is important in many areas of mathematics (though we won't use it much in this particular course) - the example of *Fourier series*. Let $C([0,1];\mathbf{C})$ be the inner product space of continuous complex-valued functions on the interval $[0,1]$, with the inner product

$$\langle f, g \rangle := \int_0^1 f(x)\overline{g(x)}\, dx.$$

Now consider the functions $\ldots, v_{-3}, v_{-2}, v_{-1}, v_0, v_1, v_2, v_3, \ldots$ in $C([0,1];\mathbf{C})$ defined by

$$v_k(x) := e^{2\pi i k x};$$

these functions are sometimes known as *complex harmonics*.

- Observe that these functions all have length 1:

$$\|v_k\| = \langle v_k, v_k \rangle^{1/2} = (\int_0^1 v_k(x)\overline{v_k(x)}\ dx)^{1/2}$$

$$= (\int_0^1 e^{2\pi ikx}e^{-2\pi ikx}\ dx)^{1/2} = (\int_0^1 1\ dx)^{1/2} = 1.$$

Also, they are all orthogonal: if $j \neq k$ then

$$\langle v_j, v_k \rangle = \int_0^1 v_j(x)\overline{v_k(x)}\ dx = \int_0^1 e^{2\pi ijx}e^{-2\pi ikx}\ dx$$

$$= \int_0^1 e^{2\pi i(j-k)x}\ dx = \frac{e^{2\pi i(j-k)x}}{2\pi i(j-k)}|_0^1 = \frac{e^{2\pi i(j-k)} - 1}{(2\pi i(j-k)}$$

$$= \frac{1-1}{2\pi i(j-k)} = 0.$$

Thus the collection $\ldots, v_{-3}, v_{-2}, v_{-1}, v_0, v_1, v_2, v_3, \ldots$ is an infinite orthonormal collection of vectors in $C([0,1]; \mathbf{C})$.

- We have not really discussed infinite bases, but it does turn out that, in some sense, that the above collection is an orthonormal basis; thus every function in $C([0,1]; \mathbf{C})$ is a linear combination of complex harmonics. (The catch is that this is an *infinite* linear combination, and one needs the theory of infinite series (as in Math 33B) to make this precise. This would take us too far afield from this course, unfortunately). This statement - which is not very intuitive at first glance - was first conjectured by Fourier, and forms the basis for something called *Fourier analysis* (which is an entire course in itself!). For now, let us work with a simpler situation.

- Define the space T_n of *trigonometric polynomials of degree at most* n to be the span of $v_0, v_1, v_2, \ldots, v_n$. In other words, T_n consists of all the functions $f \in C([0,1]; \mathbf{C})$ of the form

$$f = a_0 + a_1 e^{2\pi ix} + a_2 e^{2\pi i2x} + \ldots + a_n e^{2\pi inx}.$$

Notice that this is very similar to the space $P_n(\mathbf{R})$ of polynomials of degree at most n, since an element $f \in P_n(\mathbf{R})$ has the form

$$f = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

and so the difference between polynomials and trigonometric polynomials is that x has been replaced by $e^{2\pi ix} = \cos(2\pi x) + i\sin(2\pi x)$ (hence the name, trigonometric polynomial).

- T_n is a subspace of the inner product space $C([0,1];\mathbf{C})$, and is thus itself an inner product space. Since the vectors $v_0, v_1, \ldots, v_n$ are orthonormal, they are linearly independent, and thus $(v_0, v_1, \ldots, v_n)$ is an orthonormal basis for T_n. Thus by Theorem 6, every function f in T_n can be written as a series

$$f = a_0 + a_1 e^{2\pi ix} + a_2 e^{2\pi i2x} + \ldots + a_n e^{2\pi inx} = \sum_{j=0}^{n} a_j e^{2\pi ijx}.$$

where the (complex) scalars $a_0, a_1, \ldots, a_n$ are given by the formula

$$a_j := \langle f, v_j \rangle = \int_0^1 f(x) e^{-2\pi ijx}\, dx.$$

The coefficients a_j are known as the *Fourier coefficients* of f, and the above series is known as the *Fourier series* of f. From Corollary 3 we have the formula

$$\int_0^1 |f(x)|^2\, dx = \|f\|^2 = |a_0|^2 + |a_1|^2 + \ldots + |a_n|^2 = \sum_{j=0}^{n} |a_j|^2;$$

this is known as *Plancherel's formula.* These formulas form the foundation of Fourier analysis, and are useful in many other areas, such as signal processing, partial differential equations, and number theory. (Actually, to be truly useful, one needs to generalize these formulas to handle all kinds of functions, not just trigonometric polynomials, but to do so is beyond the scope of this course).

* * * * *

The Gram-Schmidt orthogonalization process.

- In this section all vectors are assumed to belong to a fixed inner product space V.

- In the last section we saw how many more useful properties orthonormal bases had, in comparison with ordinary bases. So it would be nice if we had some way of converting a non-orthonormal basis into an orthonormal one. Fortunately, there is such a process, and it is called *Gram-Schmidt orthogonalization.*

- To make a basis orthonormal there are really two steps; first one has to make a basis orthogonal, and then once it is orthogonal, one has to make it orthonormal. The second procedure is easier to describe than the first, so let us describe that first.

- **Definition.** A *unit vector* is any vector v of length 1 (i.e. $\|v\| = 1$, or equivalently $\langle v, v\rangle = 1$).

- **Example.** In $\mathbf{R}^2$, the vector $(3/5, 4/5)$ is a unit vector, but $(3,4)$ is not. In $C([0,1];\mathbf{C})$, the function x is not a unit vector ($\|x\|^2 = \int_0^1 xx\ dx = 1/2$), but $\sqrt{2}x$ is (why?). The 0 vector is never a unit vector. In $\mathbf{R}^3$, the vectors $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ are all unit vectors.

- Unit vectors are sometimes known as *normalized* vectors. Note that an orthogonal basis will be orthonormal if it consists entirely of unit vectors.

- Most non-zero vectors are not unit vectors, e.g. $(3,4)$ is not a unit vector. However, one can always turn a non-zero vector into a unit vector by dividing out by its length:

- **Lemma 8.** If v is a non-zero vector, then $v/\|v\|$ is a unit vector.

- **Proof.** Since v is non-zero, $\|v\|$ is non-zero, so $v/\|v\|$ is well defined. But then
$$\|v/\|v\|\| = \|\frac{1}{\|v\|}v\| = \frac{1}{\|v\|}\|v\| = 1$$
and so $v/\|v\|$ is a unit vector. □

- We sometimes call $v/\|v\|$ the *normalization* of v. If $(v_1, v_2, \ldots, v_n)$ is a basis, then we can *normalize* this basis by replacing each vector v_j by its normalization $v_j/\|v_j\|$, obtaining a new basis $(v_1/\|v_1\|, v_2/\|v_2\|, \ldots, v_n/\|v_n\|)$ which now consists entirely of unit vectors. (Why is this still a basis?)

- **Lemma 9.** If $(v_1, v_2, \ldots, v_n)$ is an orthogonal basis of an inner product space V, then the normalization $(v_1/\|v_1\|, v_2/\|v_2\|, \ldots, v_n/\|v_n\|)$ is an orthonormal basis.

- **Proof.** Since the basis $(v_1, \ldots, v_n)$ has n elements, V must be n-dimensional. Since the vectors v_j are orthogonal to each other, the vectors $v_j/\|v_j\|$ must also be orthogonal to each other (multiplying vectors by a scalar does not affect orthogonality). By Lemma 8, these vectors are also unit vectors. So the claim follows from Corollary 5. $\square$

- **Example.** The basis $((3,4), (-40,30))$ is an orthogonal basis of $\mathbf{R}^2$ (why?). If we normalize this basis we obtain $((3/5, 4/5), (-4/5, 3/5))$, and this is now an orthonormal basis of $\mathbf{R}^2$.

- So we now know how to turn an orthogonal basis into an orthonormal basis - we normalize all the vectors by dividing out their length. Now we come to the tricker part of the procedure - how to turn a non-orthogonal basis into an orthogonal one. The idea is now to subtract scalar multiples of one vector from another to make them orthogonal (you might see some analogy here with row operations of the second and third kind).

- To illustrate the idea, we first consider the problem of how to make just two vectors v, w orthogonal to each other.

- **Lemma 10.** If v and w are vectors, and w is non-zero, then the vector $v - cw$ is orthogonal to w, where the scalar c is given by the formula $c := \frac{\langle v, w\rangle}{\|w\|^2}$.

- **Proof.** We compute

$$\langle v - cw, w\rangle = \langle v, w\rangle - c\langle w, w\rangle = \langle v, w\rangle - c\|w\|^2.$$

But since $c := \frac{\langle v, w\rangle}{\|w\|^2}$, we have $\langle v - cw, w\rangle = 0$ and so $v - cw$ is orthogonal to w. $\square$

- **Example.** Let $v = (3,4)$ and $w = (5,0)$. Then v and w are not orthogonal; in fact, $\langle v, w\rangle = 15 \neq 0$. But if we replace v by the vector $v' := v - cw = (3,4) - \frac{15}{5^2}(5,0) = (3,4) - (3,0) = (0,4)$, then v' is now orthogonal to w.

- Now we suppose that we have already made k vectors orthogonal to each other, and now we work out how to make a $(k+1)^{th}$ vector also orthogonal to the first k.

- **Lemma 11.** Let $w_1, w_2, \ldots, w_k$ be orthogonal non-zero vectors, and let v be another vector. Then the vector v', defined by
$$v' := v - c_1 w_1 - c_2 w_2 - \ldots - c_k w_k$$
is orthogonal to all of $w_1, w_2, \ldots, w_k$, where the scalars $c_1, c_2, \ldots, c_k$ are given by the formula
$$c_j := \frac{\langle v, w_j \rangle}{\|w_j\|^2}$$
for all $j = 1, \ldots, k$.

- Note that Lemma 10 is just Lemma 11 applied to the special case $k = 1$. We can write v' in series notation as
$$v' := v - \sum_{j=1}^{k} \frac{\langle v, w_j \rangle}{\|w_j\|^2} w_j.$$

- **Proof.** We have to show that v' is orthogonal to each w_j. We compute
$$\langle v', w_j \rangle = \langle v, w_j \rangle - c_1 \langle w_1, w_j \rangle - c_2 \langle w_2, w_j \rangle - \ldots - c_k \langle w_k, w_j \rangle.$$
But we are assuming that the $w_1, \ldots, w_k$ are orthogonal, so all the inner products $\langle w_i, w_j \rangle$ are zero, except for $\langle w_j, w_j \rangle$, which is equal to $\|w_j\|^2$. Thus
$$\langle v', w_j \rangle = \langle v, w_j \rangle - c_j \|w_j\|^2.$$
But since $c_j := \frac{\langle v, w_j \rangle}{\|w_j\|^2}$, we thus have $\langle v', w_j \rangle$ and so v' is orthogonal to w_j. $\square$

- We can now use Lemma 11 to turn any linearly independent set of vectors into an orthogonal set.

- **Gram-Schmidt orthogonalization process.** Let $v_1, v_2, \ldots, v_n$ be a linearly independent set of vectors. Suppose we construct the vectors $w_1, \ldots, w_n$ by the formulae
$$w_1 := v_1$$

$$w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1$$

$$w_3 := v_3 - \frac{\langle v_3, w_1 \rangle}{\|w_1\|^2} w_1 - \frac{\langle v_3, w_2 \rangle}{\|w_2\|^2} w_2$$

$$\ldots$$

$$w_n := v_n - \frac{\langle v_n, w_1 \rangle}{\|w_1\|^2} w_1 - \ldots - \frac{\langle v_n, w_{n-1} \rangle}{\|w_{n-1}\|^2} w_{n-1}.$$

Then the vectors $w_1, \ldots, w_n$ are orthogonal, non-zero, and the vector space $\mathrm{span}(w_1, \ldots, w_n)$ is the same vector space as $\mathrm{span}(v_1, \ldots, v_n)$ (i.e. the vectors $w_1, \ldots, w_n$ have the same span as $v_1, \ldots, v_n$). More generally, we have that $w_1, \ldots, w_k$ has the same span as $v_1, \ldots, v_k$ for all $1 \leq k \leq n$.

- **Proof.** We prove this by induction on n. In the base case $n = 1$ we just have $w_1 := v_1$, and so clearly v_1 and w_1 has the same span. Also v_1 is an orthonormal collection of vectors (by default, since there is nobody else to be orthonormal to).

  Now suppose inductively that $n > 1$, and that we have already proven the claim for $n - 1$. In particular, we already know that the vectors $w_1, \ldots, w_{n-1}$ are orthogonal, non-zero, and that $v_1, \ldots, v_k$ has the same span as $w_1, \ldots, w_k$ for any $1 \leq k \leq n - 1$. By Lemma 11, we thus see that the vector w_n is orthogonal to $w_1, \ldots, w_{n-1}$. Now we have to show that w_n is non-zero and that $w_1, \ldots, w_n$ has the same span as $v_1, \ldots, v_n$.

  Let V denote the span of $v_1, \ldots, v_n$, and W denote the span of $w_1, \ldots, w_n$. We have to show that $V = W$. Note that W contains the span of $w_1, \ldots, w_{n-1}$, and hence contains the span of $v_1, \ldots, v_{n-1}$. In particular it contains $v_1, \ldots, v_{n-1}$, and also contains w_n. But from the formula

$$v_n = w_n + \frac{\langle v_n, w_1 \rangle}{\|w_1\|^2} w_1 + \ldots + \frac{\langle n_3, w_{n-1} \rangle}{\|w_{n-1}\|^2} w_{n-1}$$

  we thus see that W contains v_n. Thus W contains the span of $v_1, \ldots, v_n$, i.e. W contains V. But V is n-dimensional (since it is the span of n linearly independent vectors), and W is at most n dimensional (since W is also the span of n vectors), and so V and W must actually be equal. Furthermore this shows that $w_1, \ldots, w_n$ are linearly independent (otherwise W would have dimension less than n). In particular w_n is non-zero. This completes everything we need to do to finish the induction. $\square$

- **Example** Let $v_1 := (1,1,1)$, $v_2 := (1,1,0)$, and $v_3 := (1,0,0)$. The vectors v_1, v_2, v_3 are independent (in fact, they form a basis for $\mathbf{R}^3$) but are not orthogonal. To make them orthogonal, we apply the Gram-Schmidt orthogonalization process, setting

$$w_1 := v_1 = (1,1,1)$$

$$w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1 = (1,1,0) - \frac{2}{3}(1,1,1) = (\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})$$

$$w_3 := v_3 - \frac{\langle v_3, w_1 \rangle}{\|w_1\|^2} w_1 - \frac{\langle v_3, w_2 \rangle}{\|w_2\|^2} w_2$$

$$= (1,0,0) - \frac{1}{3}(1,1,1) - \frac{1/3}{6/9}(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}) = (1/2, -1/2, 0).$$

Thus we have created an orthogonal set $(1,1,1), (\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}), (\frac{1}{2}, -\frac{1}{2}, 0)$, which has the same span as v_1, v_2, v_3, i.e. it is also a basis for $\mathbf{R}^3$. Note that we can then use Lemma 8 to normalize this basis and make it orthonormal, obtaining the orthonormal basis

$$\frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{6}}(1,1,-2), \frac{1}{\sqrt{2}}(1,-1,0).$$

- We shall call the *normalized Gram-Schmidt orthogonalization process* the procedure of first applying the ordinary Gram-Schmidt orthogonalization process, and then normalizing all the vectors one obtains as a result of that process in order for them to have unit length.

- One particular consequence of Gram-Schmidt is that we always have at least one orthonormal basis lying around, at least for finite-dimensional inner product spaces.

- **Corollary 12.** Every finite-dimensional inner product space V has an orthonormal basis.

- **Proof.** Let's say V is n-dimensional. Then V has some basis $(v_1, \ldots, v_n)$. By the Gram-Schmidt orthogonalization process, we can thus create a new collection $(w_1, \ldots, w_n)$ of non-zero orthogonal vectors. By Lemma 9, we can then create a collection $(y_1, \ldots, y_n)$ of orthonormal vectors. By Corollary 5, it is an orthonormal basis of V. $\square$

* * * * *

Orthogonal complements

- We know what it means for two vectors to be orthogonal to each other, $v \perp w$; it just means that $\langle v, w \rangle = 0$. We now state what it means for two *subspaces* to be orthogonal to each other.

- **Definition.** Two subspaces V_1, V_2 of an inner product space V are said to be *orthogonal* if we have $v_1 \perp v_2$ for all $v_1 \in V_1$ and $v_2 \in V_2$, and we denote this by $V_1 \perp V_2$.

- **Example.** The subspaces $V_1 := \{(x, y, 0, 0, 0) : x, y \in \mathbf{R}\}$ and $V_2 := \{(0, 0, z, w, 0) : z, w \in \mathbf{R}\}$ of $\mathbf{R}^5$ are orthogonal, because $(x, y, 0, 0, 0) \perp (0, 0, z, w, 0)$ for all $x, y, z, w \in \mathbf{R}$. The space V_1 is similarly orthogonal to the three-dimensional space $V_3 := \{(0, 0, z, w, u) : z, w, u \in \mathbf{R}\}$. However, V_1 is not orthogonal to the one-dimensional space $V_4 := \{(t, t, t, t, t) : t \in \mathbf{R}\}$, since the inner product of $(x, y, 0, 0, 0)$ and (t, t, t, t, t) can be non-zero (e.g. take $x = y = t = 1$).

- **Example.** The zero vector space $\{0\}$ is orthogonal to any other subspace of V (why?)

- Orthogonal spaces have to be disjoint:

- **Lemma 13.** If $V_1 \perp V_2$, then $V_1 \cap V_2 = \{0\}$.

- Clearly 0 lies in $V_1 \cap V_2$ since every vector space contains 0. Now suppose for contradiction that $V_1 \cap V_2$ contained at least one other vector v, which must of course be non-zero. Then $v \in V_1$ and $v \in V_2$; since $V_1 \perp V_2$, this implies that $v \perp v$, i.e. that $\langle v, v \rangle = 0$. But this implies that $\|v\|^2 = 0$, hence $v = 0$, contradiction. Thus $V_1 \cap V_2$ does not contain any vector other than zero. $\square$

- As we can see from the above example, a subspace V_1 can be orthogonal to many other subspaces V_2. However, there is a maximal orthogonal subspace to V_1 which contains all the others:

- **Definition** The *orthogonal complement* of a subspace V_1 of an inner product space V, denoted $V_1^\perp$, is defined to be the space of all vectors perpendicular to V_1:
$$V_1^\perp := \{v \in V : v \perp w \text{ for all } w \in V_1\}.$$

- **Example.** Let $V_1 := \{(x, y, 0, 0, 0) : x, y \in \mathbf{R}\}$. Then $V_1^\perp$ is the space of all vectors $(a, b, c, d, e) \in \mathbf{R}^5$ such that (a, b, c, d, e) is perpendicular to V_1, i.e.
$$\langle (a, b, c, d, e), (x, y, 0, 0, 0) \rangle = 0 \text{ for all } x, y \in \mathbf{R}.$$
In other words,
$$ax + by = 0 \text{ for all } x, y \in \mathbf{R}.$$
This can only happen when $a = b = 0$, although c, d, e can be arbitrary. Thus we have
$$V_1^\perp = \{(0, 0, c, d, e) : c, d, e \in \mathbf{R}\},$$
i.e. $V_1^\perp$ is the space V_3 from the previous example.

- **Example.** If $\{0\}$ is the zero vector space, then $\{0\}^\perp = V$ (why?). A little trickier is that $V^\perp = \{0\}$. (Exercise! Hint: if v is perpendicular to every vector in V, then in particular it must be perpendicular to itself).

- From Lemma 1 we can check that $V_1^\perp$ is a subspace of V (exercise!), and is hence an inner product space.

- **Lemma 14.** If $V_1 \perp V_2$, then V_2 is a subspace of $V_1^\perp$. Conversely, if V_2 is a subspace of $V_1^\perp$, then $V_1 \perp V_2$.

- **Proof.** First suppose that $V_1 \perp V_2$. Then every vector v in V_2 is orthogonal to all of V_1, and hence lies in $V_1^\perp$ by definition of $V_1^\perp$. Thus V_2 is a subspace of $V_1^\perp$. Conversely, if V_2 is a subspace of $V_1^\perp$, then every vector v in V_2 is in $V_1^\perp$ and is thus orthogonal to every vector in V_1. Thus V_1 and V_2 are orthogonal. $\square$

- Sometimes it is not so easy to compute the orthogonal complement of a vector space, but the following result gives one way to do so.

- **Theorem 15.** Let W be a k-dimensional subspace of an n-dimensional inner product space V. Let $(v_1, \ldots, v_k)$ be a basis of W, and let $(v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n)$ be an extension of that basis to be a basis of V. Let $(w_1, \ldots, w_n)$ be the normalized Gram-Schmidt orthogonalization of $(v_1, \ldots, v_n)$. Then $(w_1, \ldots, w_k)$ is an orthonormal basis of W, and $(w_{k+1}, \ldots, w_n)$ is an orthonormal basis of $W^\perp$.

- **Proof.** From the Gram-Schmidt orthogonalization process, we know that $(w_1, \ldots, w_k)$ spans the same space as $(v_1, \ldots, v_k)$ - i.e. it spans W. Since W is k-dimensional, this means that $(w_1, \ldots, w_k)$ is a basis for W, which is orthonormal by the normalized Gram-Schmidt process. Similarly $(w_1, \ldots, w_n)$ spans the n-dimensional space V, which implies that it is a basis for V.

- Thus the the vectors $w_{k+1}, \ldots, w_n$ are orthonormal and thus (by Corollary 4) linearly independent. It remains to show that they span $W^\perp$. First we show that they lie in $W^\perp$. Let w_j be one of these vectors. Then w_j is orthogonal to $w_1, \ldots, w_k$, and is thus (by Lemma 1) orthogonal to their span, which is W. Thus w_j lies in $W^\perp$. In particular, the span of $w_{k+1}, \ldots, w_n$ must lie inside $W^\perp$.

- Now we show that every vector $V^\perp$ lies in the span of $w_1, \ldots, w_k$. Let v be any vector in $W^\perp$. By Theorem 6 we have

  $$v = \langle v, w_1 \rangle w_1 + \ldots + \langle v, w_n \rangle w_n.$$

  But since $v \in W^\perp$, v is orthogonal to $w_1, \ldots, w_k$, and so the first k terms on the right-hand side vanish. Thus we have

  $$v = \langle v, w_{k+1} \rangle w_{k+1} + \ldots + \langle v, w_n \rangle w_n$$

  and in particular v is in the span of $w_{k+1}, \ldots, w_n$ as desired. $\square$

- **Corollary 16 (Dimension theorem for orthogonal complements)** If W is a subspace of a finite-dimensional inner product space V, then $\dim(W) + \dim(W^\perp) = \dim(V)$.

- **Example.** Suppose we want to find the orthogonal complement of the line $W := \{(x, y) \in \mathbf{R}^2 : 3x + 4y = 0\}$ in $\mathbf{R}^2$. This example is so simple that one could do this directly, but instead we shall choose do this via Theorem 15 for sake of illustration. We first need to find a basis for W; since W is one-dimensional, we just to find one non-zero vector in W, e.g. $v_1 := (-4, 3)$, and this will be our basis. Then we extend this basis to a basis of the two-dimensional space $\mathbf{R}^2$ by adding one more linearly independent vector, for instance we could take $v_2 := (1, 0)$. This basis is not orthogonal or orthonormal, but we can apply the Gram-Schmidt process to make it orthogonal:

$$w_1 := v_1 = (-4, 3); \quad w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1 = (1, 0) - \frac{-4}{25}(-4, 3) = (\frac{9}{25}, \frac{12}{25}).$$

We can then normalize:

$$w_1' := w_1/\|w_1\| = (-\frac{4}{5}, \frac{3}{5}); \quad w_2' := w_2/\|w_2\| = (\frac{3}{5}, \frac{4}{5}).$$

Thus w_1' is an orthonormal basis for W, w_2' is an orthonormal basis for $W^\perp$, and (w_1, w_2) is an orthonormal basis for $\mathbf{R}^2$. (Note that we could skip the normalization step at the end if one only wanted an orthogonal basis for these spaces, as opposed to an orthonormal basis).

- **Example.** Let's give $P_2(\mathbf{R})$ the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\overline{g(x)}\, dx.$$

The space $P_2(\mathbf{R})$ contains $P_1(\mathbf{R})$ as a subspace. Suppose we wish to compute the orthogonal complement of $P_1(\mathbf{R})$. We begin by taking a basis of $P_1(\mathbf{R})$ - let's use the standard basis $(1, x)$, and then extend it to a basis of $P_2(\mathbf{R})$ - e.g. $(1, x, x^2)$. We then apply the Gram-Schmidt orthogonalization procedure:

$$w_1 := 1$$

$$w_2 := x - \frac{\langle x, w_1 \rangle}{\|w_1\|^2} 1 = x - \frac{0}{2} 1 = x$$

$$w_3 := x^2 - \frac{\langle x^2, w_1\rangle}{\|w_1\|^2}1 - \frac{\langle x^2, w_2\rangle}{\|w_2\|^2}x$$
$$= x^2 - \frac{2/3}{2}1 - \frac{0}{2/3}x = x^2 - 1/3.$$

We can then normalize:

$$w_1' := w_1/\|w_1\| = \frac{1}{\sqrt{2}}$$

$$w_2' := w_2/\|w_2\| = \frac{\sqrt{3}}{\sqrt{2}}x$$

$$w_3' := w_3/\|w_3\| = \frac{\sqrt{45}}{\sqrt{8}}(x^2 - \frac{1}{3}).$$

Thus $W^\perp$ has w_3' as an orthonormal basis. Or one can just use w_3 as a basis, so that

$$W^\perp = \{a(x^2 - \frac{1}{3}) : a \in \mathbf{R}\}.$$

- **Corollary 17** If W is a k-dimensional subspace of an n-dimensional inner product space V, then every vector $v \in V$ can be written in exactly one way as $w + u$, where $w \in W$ and $u \in W^\perp$.

- **Proof.** By Theorem 15, we can find a orthonormal basis $(w_1, w_2, \ldots, w_n)$ of V such that $(w_1, \ldots, w_k)$ is an orthonormal basis of W and $(w_{k+1}, \ldots, w_n)$ is an orthonormal basis of $W^\perp$. If v is a vector in V, then we can write

$$v = a_1 w_1 + \ldots + a_n w_n$$

for some scalars $a_1, \ldots, a_n$. If we write

$$w := a_1 w_1 + \ldots + a_k w_k; \quad u := a_{k+1} w_{k+1} + \ldots + a_n w_n$$

then we have $w \in W, u \in W^\perp$, and $v = w + u$. Now we show that this is the only way to decompose v in this manner. If $v = w' + u'$ for some $w' \in W$, $u' \in W^\perp$, then

$$w + u = w' + u'$$

and so

$$w - w' = u' - u.$$

But $w - w'$ lies in W, and $u' - u$ lies in $W^\perp$. By Lemma 13, this vector must be 0, so that $w = w'$ and $u = u'$. Thus there is no other way to write $v = w + u$. $\square$

- We call the vector w obtained in the above manner the *orthogonal projection* of v onto W; this terminology can be made clear by a picture (at least in $\mathbf{R}^3$ or $\mathbf{R}^2$), since w, u, and v form a right-angled triangle whose base w lies in W. This projection can be computed as

$$w = a_1 w_1 + \ldots + a_k w_k = \langle v, w_1 \rangle w_1 + \ldots + \langle v, w_k \rangle w_k$$

where $w_1, \ldots, w_k$ is any orthonormal basis of W.

- **Example** Let $W := \{(x, y) \in \mathbf{R}^2 : 3x + 4y = 0\}$ be as before. Suppose we wish to find the orthogonal projection of the vector $(1, 1)$ to W. Since we have an orthonormal basis given by $w_1' := (-\frac{4}{5}, \frac{3}{5})$, we can compute the orthogonal projection as

$$w = \langle (1, 1), w_1' \rangle w_1' = -\frac{1}{5}(-\frac{4}{5}, \frac{3}{5}) = (\frac{4}{25}, -\frac{3}{25}).$$

- The orthogonal projection has the "nearest neighbour" property:

- **Theorem 18.** Let W be a subspace of a finite-dimensional inner product space V, let v be a vector in V, and let w be the orthogonal projection of v onto W. Then w is closer to v than any other element of W; more precisely, we have $\|v - w'\| > \|v - w\|$ for all vectors w' in W other than w.

- **Proof.** Write $v = w + u$, where $w \in W$ is the orthogonal projection of v onto W, and $u \in W^\perp$. Then we have $\|v - w\| = \|u\|$. On the other hand, to compute $v - w'$, we write $v - w' = (v - w) + (w - w') = u + (w - w')$. Since w, w' lie in W, $w - w'$ does also. But u lies in $W^\perp$, thus $u \perp w - w'$. By Pythagoras's theorem we thus have

$$\|v - w'\|^2 = \|u\|^2 + \|w - w'\|^2 > \|u\|^2$$

(since $w \neq w'$) and so $\|v - w'\| > \|u\| = \|v - w\|$ as desired. $\square$

- This Theorem makes the orthogonal projection useful for approximating a given vector v in V by a another vector in a given subspace W of V.

- **Example** Consider the vector x^2 in $P_2(\mathbf{R})$. Suppose we want to find the linear polynomial $ax + b \in P_1(\mathbf{R})$ which is closest to x^2 (using the inner product on $[-1, 1]$ from the previous example to define length). By Theorem 18, this linear polynomial will be the orthogonal projection of x^2 to $P_1(\mathbf{R})$. Using the orthonormal basis $w_1' = \frac{1}{\sqrt{2}}$, $w_2' = \frac{\sqrt{3}}{\sqrt{2}}x$ from the prior example, we thus see that this linear polynomial is

$$\langle x^2, w_1' \rangle w_1' + \langle x^2, w_2' \rangle w_2' = \frac{2/3}{\sqrt{2}} \frac{1}{\sqrt{2}} + 0 \frac{\sqrt{3}}{\sqrt{2}} x = \frac{1}{3} + 0x.$$

Thus the function $1/3 + 0x$ is the closest linear polynomial to x^2 using the inner product on $[-1, 1]$. (If one uses a different inner product, one can get a different "closest approximation"; the notion of closeness depends very much on how one measures length).