

The Gram-Schmidt Process

How and Why it Works

This is intended as a complement to §5.4 in our textbook. I assume you have read that section, so I will not repeat the definitions it gives. Our goal is to work out what happens in the proof of Theorem 5.6. We start with a finite dimensional space W with an inner product (\vec{u}, \vec{v}) . The book's statement of the theorem is slightly more complex, mentioning one vector space V that has an inner product and a subspace W that is finite dimensional, but the only role V really plays is to guarantee that W has an inner product, so we will just jump to that assumption.

Stated abstractly, what we get is this version of Theorem 5.6:

If W is a vector space of dimension m , with $m > 0$ (i.e. W is not just $\vec{0}$), and there is an inner product (\cdot, \cdot) on W , then there is an orthonormal basis $T = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$ for W .

But the critical part for our purposes is how we find T , not just that there is such a T . The method of finding T is the proof of the theorem, since it always works (given the assumptions in the statement of the theorem) and will produce T . And the method for finding T is a useful tool: An orthonormal basis for a vector space is a nice thing to have for several reasons, not the least of which is the ease of finding coordinates with respect to such a basis.

We start with some basis for W : This must exist and must contain m vectors, since W has dimension $m > 0$. We give it a name, and also name the vectors making up this starting basis, as $S = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m\}$. We will “build” the orthonormal basis T one step at a time, using the vectors \vec{u}_j in this basis. For our new basis W we need to produce vectors \vec{w}_j which (a) have magnitude 1, i.e. $\|\vec{w}_j\| = 1$ for each j , and (b) are orthogonal, i.e. $(\vec{w}_i, \vec{w}_j) = 0$ if $i \neq j$. (Of course if W were itself orthonormal, there would be no point in changing it. And indeed the process we carry out would not result in any change to W in that case, but it might be quite a bit of wasted calculation!)

We build the new basis T in two big steps. First we create an intermediate basis $\{\vec{v}_1, \dots, \vec{v}_m\}$ which is orthogonal, but these vectors might not have magnitude 1, and later we worry about the magnitude.

The way we create this orthogonal basis works can be pictured geometrically, for the first couple of steps. The picture assumes the vector space is \mathbb{R}^2 , at first, and then \mathbb{R}^3 , but the picture's purpose is just to suggest to us the right things to do and we actually carry them out in whatever the space W is, using the operations (vector addition, scalar multiplication, and inner product) W is equipped with.

We let the first vector \vec{v}_1 be the first vector from S , i.e. $\vec{v}_1 = \vec{u}_1$. Since we are not yet worrying about normality, getting the norm of the vector to be 1, that will be fine so long as $\vec{u}_1 \neq \vec{0}$: If \vec{u}_1 were $\vec{0}$, it could not be part of a basis. So the fact that S is a basis guarantees that $\vec{u}_1 \neq \vec{0}$. If $\dim(W) = m = 1$, we are now have enough “orthogonal” vectors (what orthogonal might mean with only one vector need not be worried about) since any basis for W will contain just one vector. So in that case we can skip on down to making the vector in the basis have norm 1. But if $m = \dim(W) > 1$, we need to go on and find a second vector, which we will call \vec{v}_2 , that is orthogonal to \vec{v}_1 .

In Figure 5.19, page 321 in the textbook, we see how we do this in the plane. Using language from Math 222/234, we find the component of \vec{u}_2 that is perpendicular to \vec{v}_1 : We calculate the projection of \vec{u}_2 onto \vec{v}_1 and subtract that from \vec{u}_2 . In the picture, first think of the solid arrow labelled \vec{v}_2 as just denoting the direction perpendicular to \vec{u}_2 . \vec{u}_2 can be written as the sum of two vectors, one parallel to \vec{u}_1 and one perpendicular to \vec{u}_1 : If you take the dashed arrow as the component perpendicular to \vec{u}_1 , you see that subtracting the projection of \vec{u}_2 onto \vec{v}_1 , which is the component parallel to \vec{u}_1 , from \vec{u}_2 , “slides” the dashed arrow back over to the solid \vec{v}_2 .

Inspired by that action in \mathbb{R}^2 , in our space W we emulate it. The projection of \vec{u}_2 onto \vec{v}_1 is (written in the lower left part of Figure 5.19) $\frac{(\vec{u}_2, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1$. Back in 222/234 that inner product was the dot product, but the effect is the same. We subtract that projection from \vec{u}_2 and get $\vec{v}_2 = \vec{u}_2 - \frac{(\vec{u}_2, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1$.

Our picture for the case of \mathbb{R}^2 suggests this choice for \vec{v}_2 should work, but does it? We need to know that (a) this \vec{v}_2 really is orthogonal to \vec{v}_1 , and (b) that it is not $\vec{0}$. (We will be using Theorem 5.4 from §5.3 to show linear independence for our basis: That assumed that we had an orthogonal set of vectors but it also specifically needed that the vectors were non-zero. In addition we will want to see that each $\vec{v}_i \neq \vec{0}$ for internal purposes in our calculations, as we construct it.) To check (a) we just compute the inner product (\vec{v}_1, \vec{v}_2) : If that gives 0 (the number, not the vector!) that will tell us the vectors are orthogonal. We have $(\vec{v}_1, \vec{v}_2) = (\vec{v}_1, \vec{u}_2 - \frac{(\vec{u}_2, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1)$. The properties $(\vec{u}, \vec{v} + \vec{w}) = (\vec{u}, \vec{v}) + (\vec{u}, \vec{w})$ and $(\vec{u}, c\vec{v}) = c(\vec{u}, \vec{v})$ are exactly what we need to expand that (sort of multiply it out) as $(\vec{v}_1, \vec{v}_2) = (\vec{v}_1, \vec{u}_2) - \frac{(\vec{u}_2, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} (\vec{v}_1, \vec{v}_1) = (\vec{v}_1, \vec{u}_2) - \frac{(\vec{v}_1, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} (\vec{u}_2, \vec{v}_1)$. The fraction $\frac{(\vec{v}_1, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}$ amounts to $\frac{\|\vec{v}_1\|^2}{\|\vec{v}_1\|^2}$: Since $\vec{v}_1 \neq \vec{0}$, we know $\|\vec{v}_1\| \neq 0$, so the fraction is 1 and we have $(\vec{v}_1, \vec{v}_2) = (\vec{v}_1, \vec{u}_2) - (\vec{u}_2, \vec{v}_1)$. We know that for any inner product, for any vectors \vec{x} and \vec{y} , $(\vec{x}, \vec{y}) = (\vec{y}, \vec{x})$, so that difference works out to 0. So now we have that \vec{v}_1 and \vec{v}_2 are orthogonal, part (a) above.

As we saw we needed $\vec{v}_1 \neq \vec{0}$, and in the succeeding steps we will need $\vec{v}_2 \neq \vec{0}$ in order to get our third vector: I.e. we still need to check (b). Suppose $\vec{v}_2 = \vec{0}$. From the way we produced \vec{v}_2 that would tell us $\vec{u}_2 - \frac{(\vec{u}_2, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1 = \vec{0}$. Then we would have that \vec{u}_2 was a number (the fraction) times \vec{v}_1 , and $\vec{v}_1 = \vec{u}_1$, so \vec{u}_2 would be a multiple of \vec{u}_1 , i.e. \vec{u}_2 would be a linear combination of preceding vectors in S , but S was a basis, hence linearly independent, so that cannot happen and we know \vec{v}_2 cannot be $\vec{0}$.

Again we are at a decision point: If $m = 2$, we have enough orthogonal vectors so we go to the step where we make the norms 1. But if $m > 2$ we need to keep going and find a third vector \vec{v}_3 that is orthogonal to both the two we have so far, \vec{v}_1 and \vec{v}_2 . This will be the last time we will be able to use a picture to motivate our choice, since it is hard to draw 4-dimensional pictures. That makes it critical to think as we go about creating \vec{v}_3 just what is happening, in a way that can be generalized to $m > 3$ after we get \vec{v}_3 . Figure 5.20 on page 322 in the textbook attempts to show us what to do in \mathbb{R}^3 , but since the picture is flat in the book it takes careful examination to see what is going on. The grey area at the bottom is supposed to be a copy of 2-dimensional space, the plane, representing the span of the two vectors \vec{v}_1 and \vec{v}_2 . To me the picture makes it look like \vec{v}_2 sticks up a little out of that “floor”: It is supposed to be lying in the floor. \vec{v}_3 , the vector pointing straight up, is the thing we want to find: Something orthogonal to each of \vec{v}_1 and \vec{v}_2 . Since we have $m > 3$ (or we would not get to this stage), there is a third vector \vec{u}_3 in the basis S we are using as raw materials to build our new basis. The picture suggests that \vec{u}_3 is not down in the grey plane, the span of \vec{v}_1 and \vec{v}_2 . Is that really so, and why must it be? \vec{v}_1 was the same as \vec{u}_1 , and \vec{v}_2 was a linear combination of \vec{u}_2 and \vec{v}_1 . So if \vec{u}_3 were in the span of \vec{v}_1 and \vec{v}_2 it would be a linear combination of them, hence of \vec{u}_1 and \vec{u}_2 , but then the basis S would not be linearly independent. So the picture does tell the truth, \vec{u}_3 cannot be in the span of \vec{v}_1 and \vec{v}_2 .

So \vec{u}_3 really does stick out of the grey plane. We want to subtract from it some projection into the grey plane, the dashed vector in the picture, to get a vector perpendicular to the plane and hence orthogonal to \vec{v}_1 and to \vec{v}_2 . A vector in the plane would be some linear combination of \vec{v}_1 and \vec{v}_2 , and what appears to be the right combination is found by using the projections of \vec{u}_3 onto those two vectors \vec{v}_1 and \vec{v}_2 . In the picture the dashed vector is labelled $\frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1 + \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)} \vec{v}_2$. The first part of that is the projection of \vec{u}_3 onto \vec{v}_1 and the second part is the projection of \vec{u}_3 onto \vec{v}_2 . The picture suggests that subtracting that from \vec{u}_3 will give us the orthogonal vector \vec{v}_3 that we are looking for. But to be sure it works, we have to mimic (a) and (b) above: Is it true that this

gives us a vector orthogonal to \vec{v}_1 ? We check that by computing the inner product with \vec{v}_1 . Does it give a vector orthogonal to \vec{v}_2 ? Again an inner product, this time with \vec{v}_2 , will be zero if and only if that is so. And we also want to check that this gives us a non-zero vector. Note that these checks would really be necessary even in \mathbb{R}^3 , since the picture could be misleading, but in a general vector space where we don't have a picture these tests are the real essence of the process: The pictures just suggested what might work. Even if you had just guessed that this way of constructing \vec{v}_3 might be useful, the calculations showing orthogonality and non-zero are the only things available to us for a vector space in general.

The text does carry out these tests. But it introduces new letters (the b_1, b_2, b_3 at the top of page 322): They make the calculations easier to write down, but may make it hard to realize that all they are doing is to compute the inner products mentioned in the previous paragraph. If the version there “works for you”, fine. But I will work one of them out in the direct but messy way in case that makes it easier to see what is really happening.

We choose our \vec{v}_3 in the way described, i.e. we subtract from \vec{u}_3 the sum of the projections $\frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}\vec{v}_1 + \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)}\vec{v}_2$. So we have

$$\vec{v}_3 = \vec{u}_3 - \frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}\vec{v}_1 - \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)}\vec{v}_2.$$

We need to show that choice of \vec{v}_3 “works”, i.e. (a) it is orthogonal to both \vec{v}_1 and \vec{v}_2 , and (b) that it is not $\vec{0}$.

For (a) we compute both (\vec{v}_1, \vec{v}_3) and (\vec{v}_2, \vec{v}_3) , and hope both turn out to be 0. I will just do one, the other works the same way, but concentrate on what is really happening so that we can see why it would also work when we get to \vec{v}_4, \vec{v}_5 , etc. replacing \vec{v}_3 . We work out $(\vec{v}_1, \vec{v}_3) = (\vec{v}_1, \vec{u}_3 - \frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}\vec{v}_1 - \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)}\vec{v}_2)$. As before we can “multiply out” and pull any scalars out front, getting $(\vec{v}_1, \vec{u}_3) - \frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}(\vec{v}_1, \vec{v}_1) - \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)}(\vec{v}_1, \vec{v}_2)$. Since we already knew $\vec{v}_1 \neq \vec{0}$, the $\frac{(\vec{v}_1, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}$ in the second term amounts to 1, so the second term reduces to $-(\vec{u}_3, \vec{v}_1)$. We also already knew that \vec{v}_1 and \vec{v}_2 were orthogonal, so (\vec{v}_1, \vec{v}_2) in the third term is zero. So that whole expression simplifies to $(\vec{v}_1, \vec{u}_3) - (\vec{u}_3, \vec{v}_1)$ which is 0, and we have (a).

Now for (b): We chose $\vec{v}_3 = \vec{u}_3 - \frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)}\vec{v}_1 - \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)}\vec{v}_2$: If that gave $\vec{0}$, then we could move \vec{u}_3 to the other side and see that it would be a linear combination of the preceding \vec{v}_i 's which in turn were linear combinations of earlier \vec{u}_i 's, so \vec{u}_3 would be a linear combination of preceding \vec{u}_i 's, contradicting the linear independence of the basis S . So \vec{v}_3 cannot be $\vec{0}$.

Now we want to say “keep doing this, producing vectors \vec{v}_i for $i = 1, 2, \dots, m$, such that (a) each \vec{v}_i is orthogonal to those that came before it and (b) each is also non-zero.” To be sure we can say that we need to make sure we understand how the repetitive process works.

When we have constructed $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_i$, and want to construct the next one (which we have done above for $i = 1, 2, 3$), we first define \vec{v}_{i+1} to be \vec{u}_{i+1} minus the sum of the projections of \vec{u}_{i+1} onto all of the earlier \vec{v}_j 's. Those projections are, for each j from 1 to i , $\frac{(\vec{u}_{i+1}, \vec{v}_j)}{(\vec{v}_j, \vec{v}_j)}\vec{v}_j$. As we constructed each of those earlier \vec{v}_j 's, we made sure $\vec{v}_j \neq \vec{0}$, so the fraction does not have a zero denominator, i.e. we can do this. When we want to show (a) that our new \vec{v}_i is orthogonal to each of the ones previously produced, we take the inner product of

$$\vec{v}_{i+1} = \vec{u}_{i+1} - \left[\text{the sum of } \frac{(\vec{u}_{i+1}, \vec{v}_k)}{(\vec{v}_k, \vec{v}_k)}\vec{v}_k \text{ where the sum runs over } k \text{ from } 1 \text{ to } i \right]$$

with \vec{v}_j for each $j \leq i$. “Multiplying out” that inner product we get, for any particular j ,

$$(\vec{v}_j, \vec{u}_{i+1}) - \left[\text{the sum of } \frac{(\vec{u}_{i+1}, \vec{v}_k)}{(\vec{v}_k, \vec{v}_k)}(\vec{v}_j, \vec{v}_k) \text{ where the sum runs over } k \text{ from } 1 \text{ to } i \right]$$

Now in that sum, since j is some number between 1 and i , one of the terms has $k = j$ and the rest have $k \neq j$. Whenever $k \neq j$, the factor (\vec{v}_j, \vec{v}_k) will be zero, since we already had that each of the earlier \vec{v}_j 's were orthogonal. For the one term where $j = k$, that part of the sum is $-\frac{(\vec{u}_{i+1}, \vec{v}_k)}{(\vec{v}_k, \vec{v}_k)}(\vec{v}_k, \vec{v}_k)$, which is equal to $-(\vec{u}_{i+1}, \vec{v}_k)$ since (\vec{v}_k, \vec{v}_k) being nonzero) the fraction $\frac{(\vec{v}_k, \vec{v}_k)}{(\vec{v}_k, \vec{v}_k)}$ gives 1. So the inner product reduces to $(\vec{v}_j, \vec{u}_{i+1}) - (\vec{u}_{i+1}, \vec{v}_j)$ which is zero.

The reason that our new \vec{v}_{i+1} is not $\vec{0}$ is just like we saw in the first cases: By the way we constructed \vec{v}_{i+1} , if it were $\vec{0}$, then \vec{u}_{i+1} could be written as a linear combination of earlier vectors \vec{u}_j , contradicting the linear independence of the basis S .

We can keep doing this as long as there are any more vectors \vec{u}_{i+1} in the basis S from which to work, i.e. the process stops when we have produced m new vectors \vec{v}_i : Our set $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$ are orthogonal and non-zero, so we know they are linearly independent, and our space had dimension m so they must be a basis. But we are not through: An orthonormal basis requires they each have norm 1 also. So we now replace each \vec{v}_i by $\frac{1}{\|\vec{v}_i\|}\vec{v}_i$, which we can do since $\|\vec{v}_i\| \neq 0$. The result has norm 1, but did it mess up the orthogonality we worked so hard to get? (Intuitively it would seem OK, since multiplying by that positive constant should not change the direction and orthogonality was supposed to correspond to being perpendicular. But since who knows what vector space this is all happening in, reasoning from intuition is hazardous at best!) We have replaced \vec{v}_i by a constant times \vec{v}_i : If we compute the inner product for our new vectors, (\vec{v}_i, \vec{v}_j) is $(c_1\vec{v}_i, c_2\vec{v}_j)$ for our earlier vectors \vec{v}_i, \vec{v}_j and for some constants c_1 and c_2 . But that will be $c_1c_2(\vec{v}_1, \vec{v}_2)$ for the vectors \vec{v} that we produced, that were orthogonal, so that is $c_1c_2 \times 0 = 0$ and our new, now magnitude 1, vectors are still orthogonal.

Some Examples

OK, we (I hope) now see why this process should work. But surely some examples would help really to understand what all of that was talking about!

Almost a Real Example:

For a first example I will start with some vectors in \mathbb{R}^3 . Here the inner product will be the dot product you may also have used in physics, engineering, etc., and I will also try to suggest a reason you might want to do this in geology. If you just want to see the G-S process at work numerically, skip the next few paragraphs. But I think some might like to see how this could be useful in a real problem, which requires our getting our hands dirty with some stuff about coordinates for the real world.

Let me start by emphasizing I am not a geologist, but it seems to me this example might be comprehensible by people not in geology and give some insight into why the Gram-Schmidt process could be useful: Whatever your interests are, try to think of some parallel (no pun intended!) instance in your area. Suppose there is an underground “fault” system, that I will simplify to two regions of stone that want to slide along each other but are held back by friction. (Friction in a very general sense: For example, there can be big rocks that lie across the boundary, trying to pin the two sides together so that they cannot slide.) Your measuring instruments, your maps, maybe your internal view of the world (many of us automatically orient the external world around a North-South line, for example), all have built in a set of axes which I will assume is NS (North-South) for the x -axis, EW for the y -axis, and “up” for the z -axis. So when you get data from your instruments they are reported in those coordinates. (Anybody who accuses me of being realistic has already ignored many simplifications I have made. For example, NS is not really parallel to the axis on which the earth turns, except at the equator, and instruments that make their reports relative to GPS data are probably referring to “real” NS, not NS along the tangent plane to the earth where I am!)

But for modelling the actual interactions at the interface between the material on two sides

of the fault, we would like to think of forces parallel to the interface and forces perpendicular to it. (Another, terribly gross, simplification is that I will act as if the fault lies in a plane, that it is completely flat.) Why would we want to do that? As a first approximation, the forces perpendicular to the fault are the ones determining how much the fault resists slipping, while forces along the fault are the ones making it want to slip.

We have a standard basis for \mathbb{R}^3 that we like to use, $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ in one form

or \vec{i} and \vec{j} and \vec{k} in another. One of the reasons we like that basis is that it is orthonormal: That is why the coordinates of any vector are given by its dot product with the basis vectors, making coordinates easy to compute. But from a physical standpoint, the “ortho” part has an additional significance: A force (or other physical effect) along one axis has no effect in the directions of the other axes. They are “independent” in many of their physical characteristics. What we need now is another basis, nicely oriented with respect to that fault plane in the earth, but we would still like it to be orthogonal and if possible normalized. The Gram-Schmidt process gives us exactly that.

We need to have some basis to start with. We don’t want the standard basis, using G-S on that would just give us back that same basis since that was already orthonormal. Instead we want to get one of the basis vectors in our resulting basis to be perpendicular to the fault plane. In our simplified model where the fault really is planar, if one vector in an orthogonal set is perpendicular to the plane then the others will be forced to be in the plane, so they will be set up for representing the slip forces and the first one would be good for representing the resistance forces. To go further, we have “know” where the fault plane is located. I will take the NS axis to be the x -axis, with x positive in the North direction. I will take the y axis to be East-West, with positive y being in the West direction: That way if I look down on the axes from above I see x and y oriented the way we have been drawing them in math classes, so that with z positive upward we would have cross products working the way they did in Math 234. (If we made y go the other direction, cross products would have the opposite signs.)

If you wanted to skip all the stuff about real-world orientation, here is where you should start in again. For an example, and to make the calculations simple, I will assume that the fault plane

is tipped up out of the horizontal plane in such a way that the vector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is perpendicular to

the fault. So I want one of our eventual basis vectors to be in that direction. If we remember that the first vector \vec{v}_1 that we produced was just the first vector from our starting basis, it seems useful

to let that vector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ be a basis vector in the set (called S earlier) that we start with. I don’t

for the moment care which direction the other two vectors, in the fault plane, go, so I just need to find two more vectors that can go with this one to give a basis. For simplicity I will pick the north-pointing vector and the west-pointing vector, giving as my starting basis S the set

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

(If there were a reason to prefer an orientation in the fault plane, e.g. one basis vector horizontal so that the third is more or less vertical and could be used easily in representing water flow that influences slippage, one could choose specific vectors for the second and/or third basis vectors here. But that would be too realistic for a math class...)

That is a basis for \mathbb{R}^3 , that needs to be checked but is pretty easy. So we start with that and from here on out we just apply the Gram-Schmidt process. To use the same notation as in the section where we described the process, we have $\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, and $\vec{u}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. So our

first step is just the renaming, $\vec{v}_1 = \vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Next we compute our second orthogonal vector \vec{v}_2

using the scheme above: Calculationally it is $\vec{v}_2 = \vec{u}_2 - \frac{(\vec{u}_2, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1$, and geometrically it is \vec{u}_2 minus the

projection of \vec{u}_2 on $\vec{v}_1 = \vec{u}_1$. Carrying that out with $\vec{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, and remembering

that our inner product is the dot product, we have $(\vec{u}_2, \vec{v}_1) = 1 \times 1 + 0 \times 1 + 0 \times 1 = 1$. (As I said, the numbers were chosen to make calculations pretty easy. I won't keep writing out how the dot product is calculated, this was just a reminder of how we multiply corresponding entries and add up the results.) Similarly $(\vec{v}_1, \vec{v}_1) = 3$. So

$$\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}.$$

Now that was not too painful, was it! (If you are the suspicious type, compute $(\vec{v}_1, \vec{v}_2) = \frac{2}{3} - \frac{1}{3} - \frac{1}{3} = 0$ to confirm that those two vectors are orthogonal.)

We go on to construct the third orthogonal basis vector, \vec{v}_3 , using the formula from above $\vec{v}_3 = \vec{u}_3 - \frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1 - \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)} \vec{v}_2$. First I will compute the inner products (dot products in this case) to use in the calculation: $(\vec{u}_3, \vec{v}_1) = 1$, (\vec{v}_1, \vec{v}_1) is still 3, $(\vec{u}_3, \vec{v}_2) = -\frac{1}{3}$, and $(\vec{v}_2, \vec{v}_2) = \frac{4}{9} + \frac{1}{9} + \frac{1}{9} = \frac{6}{9}$.

$$\begin{aligned} \text{So } \vec{v}_3 &= \vec{u}_3 - \frac{1}{3} \vec{v}_1 - \frac{-\frac{1}{3}}{\frac{6}{9}} \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} + \begin{bmatrix} \frac{2}{6} \\ -\frac{1}{6} \\ -\frac{1}{6} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}. \end{aligned}$$

Now it is easy to check that each pair of vectors is orthogonal by taking the inner products and getting zero. So this,

$$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}, \begin{bmatrix} 0 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \right\},$$

is our orthogonal basis, we just need to make the magnitudes all 1. The present magnitude of \vec{v}_1 is $\sqrt{1+1+1} = \sqrt{3}$, so we replace it by $\frac{1}{\sqrt{3}} \vec{v}_1$ which we could write that way or in several other forms

such as $\begin{bmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}$. For \vec{v}_2 the magnitude so far is $\sqrt{\frac{4}{9} + \frac{1}{9} + \frac{1}{9}} = \frac{\sqrt{6}}{3}$, so we use for our second basis

vector, with magnitude 1, the vector \vec{v}_2 that we had, multiplied by $\frac{1}{\frac{\sqrt{6}}{3}} = \frac{3}{\sqrt{6}}$. Depending on how seriously your high-school algebra teacher threatened you about always rationalizing denominators,

there are several ways this could be written, e.g. $\begin{bmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix}$. Now we do the same for \vec{v}_3 , which has

magnitude $\sqrt{0 + \frac{1}{4} + \frac{1}{4}} = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}}$, i.e. we divide by that fraction which amounts to multiplying

by $\sqrt{2}$, and we get $\begin{bmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$ as our third basis vector, so the orthonormal basis we get is

$$\left\{ \begin{bmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{bmatrix}, \begin{bmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix}, \begin{bmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} \right\} \text{ which you can check is both "ortho" and "normal".}$$

Another Example:

This example does not pretend to look like it has to do with the real world. (But in fact it does: What it produces are (except for the way they are normalized) what are called the Legendre Polynomials, which are used extensively in physical sciences and economics. We just don't say anything about how they might be used.)

We start with the vector space P_4 , the set of polynomials of degree at most 4 with the usual addition and scalar multiplication rules. On that space we introduce an inner product as follows: If $p(t)$ and $q(t)$ are vectors, i.e. polynomials of degree at most 4, we think of them as functions and for the inner product (p, q) we compute the integral $\int_{-1}^1 p(t) \times q(t) dt$. As shown in class and in the book, that does provide an inner product meeting the requirements of the definition.

Our "nice" basis for P_4 has been $\{1, t, t^2, t^3\}$. But that is neither orthogonal nor normal. Call those four vectors $\vec{u}_1 \dots \vec{u}_4$. Then $(1, t^2) = \int_{-1}^1 1 \times t^2 dt = \frac{2}{3}$, not 0, so \vec{u}_1 and \vec{u}_3 are not orthogonal.

And $(1, 1) = \int_{-1}^1 1 \times 1 dt = 2$, so $\|\vec{u}_1\| = \sqrt{2} \neq 1$. But we can apply the Gram-Schmidt process to that basis and get an orthonormal basis for P_4 .

We start by setting $\vec{v}_1 = \vec{u}_1 = 1$. We want to let $\vec{v}_2 = \vec{u}_2 - \frac{(\vec{u}_2, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1$, so we compute $(\vec{u}_2, \vec{v}_1) = \int_{-1}^1 t \times 1 dt = 0$. What happened??? The basis we started with, while not all pairs of vectors were orthogonal, did have \vec{u}_1 and \vec{u}_2 orthogonal! So when we compute \vec{v}_2 we just get $\vec{u}_2 - \vec{0} = \vec{u}_2 = t$.

Next we compute $\vec{v}_3 = \vec{u}_3 - \frac{(\vec{u}_3, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1 - \frac{(\vec{u}_3, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)} \vec{v}_2$. To do this we use $(\vec{u}_3, \vec{v}_1) = \int_{-1}^1 t^2 \times 1 dt = \frac{2}{3}$, $(\vec{v}_1, \vec{v}_1) = \int_{-1}^1 1 \times 1 dt = 2$ as computed above, $(\vec{u}_3, \vec{v}_2) = \int_{-1}^1 t^2 \times t dt = 0$ (another pair already orthogonal), and $(\vec{v}_2, \vec{v}_2) = \int_{-1}^1 t \times t dt = \frac{2}{3}$. So $\vec{v}_3 = \vec{u}_3 - \frac{2/3}{2} \vec{v}_1 - \frac{0}{2/3} \vec{v}_2 = t^2 - \frac{1}{3} \times 1 - 0 = t^2 - \frac{1}{3}$.

Proceeding to $\vec{v}_4 = \vec{u}_4 - \frac{(\vec{u}_4, \vec{v}_1)}{(\vec{v}_1, \vec{v}_1)} \vec{v}_1 - \frac{(\vec{u}_4, \vec{v}_2)}{(\vec{v}_2, \vec{v}_2)} \vec{v}_2 - \frac{(\vec{u}_4, \vec{v}_3)}{(\vec{v}_3, \vec{v}_3)} \vec{v}_3$: We already computed (\vec{v}_1, \vec{v}_1) and got 2, and (\vec{v}_2, \vec{v}_2) was $\frac{2}{3}$. The new things we need are: $(\vec{u}_4, \vec{v}_1) = \int_{-1}^1 t^3 \times 1 dt = 0$; $(\vec{u}_4, \vec{v}_2) =$

$\int_{-1}^1 t^3 \times t dt = \frac{2}{5}$; $(\vec{u}_4, \vec{v}_3) = \int_{-1}^1 t^3 \times t^2 - \frac{1}{3} dt = 0$; and $(\vec{v}_3, \vec{v}_3) = \int_{-1}^1 t^3 \times t^2 - \frac{1}{3} dt = \frac{8}{45}$. So $\vec{v}_4 = \vec{u}_4 - 0 \times \vec{v}_1 - \frac{3}{5} \times \vec{v}_2 - 0 \times \vec{v}_3 = t^3 - \frac{3}{5}t$.

Now we have as an orthonormal basis $\{1, t, t^2 - \frac{1}{3}, t^3 - \frac{3}{5}t\}$. To make the magnitudes 1 we multiply each by $1/(\text{its current magnitude})$. The magnitude of $p(t)$ is

$$\|p(t)\| = \sqrt{(p(t), p(t))} = \sqrt{\int_{-1}^1 p(t) \times p(t) dt}.$$

For 1 we already computed $(1, 1) = 2$, so the first vector in our orthonormal basis is $\frac{1}{\sqrt{2}} \times 1 = \frac{\sqrt{2}}{2}$. The magnitude of t is $\sqrt{(t, t)}$ and (t, t) was already computed as $\frac{2}{3}$, so the second vector is $\sqrt{\frac{3}{2}} \times t = \frac{\sqrt{6}}{2}t$. For the third vector, $\|t^3 - \frac{3}{5}\|^2 = \int_{-1}^1 (t^3 - \frac{3}{5})^2 dt = \frac{8}{45}$, so the third vector as normalized is $\sqrt{\frac{45}{8}} \times (t^3 - \frac{3}{5}) = \frac{3\sqrt{10}}{4}t^2 - \frac{\sqrt{10}}{4}$. Finally, our fourth vector had magnitude $\sqrt{\int_{-1}^1 (t^3 - \frac{3}{5}t)^2 dt} = \sqrt{\frac{8}{175}} = \frac{2\sqrt{14}}{35}$, giving $\frac{5}{4}\sqrt{14}t^3 - \frac{3}{4}\sqrt{14}t$ as the fourth vector in our orthonormal basis,

$$\left\{ \frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{2} t, \frac{3\sqrt{10}}{4} t^2 - \frac{\sqrt{10}}{4}, \frac{5}{4}\sqrt{14} t^3 - \frac{3}{4}\sqrt{14} t \right\}.$$

Real World Ramifications

The Gram-Schmidt process is something that is actually useful in the real world. I won't go into how it might be used in different application areas, but I do want to point out a couple of things about how it would be carried out in reality. The first thing to realize is that "arithmetic" in the real world is likely to be done on a computer, and computers don't do perfect arithmetic. Except with very special programming languages, they only carry along some fixed number of digits. In many modern languages you get a choice, for numbers that are allowed to be very large or small and hence cannot be stored just as integers, of formats such as "real" or "float", "double", and maybe "quad". See the Wikipedia article

http://en.wikipedia.org/wiki/Floating_point

for a discussion of how much precision those actually provide: The point is that there will be some fixed number of digits available. That number may seem large, compared to the accuracy of whatever data you might be working from, but as calculations are done on the data over and over, errors resulting from the finite precision build up. Consider the following calculation:

$$w = \frac{x}{y - z}.$$

If the values of x , y , and z are known to, for example, 10 digits, if y and z were the same for the first nine digits, then $y - z$ might both be small and also not have much precision at all. Then dividing that small quantity into x might produce a very large number w , with little precision. (The size is not so relevant as the loss of precision. But any of those number formats also has a largest number it can represent, and a really bad kind of error occurs if w exceeds that bound.) And then if w is used in further calculations, very soon the errors can overwhelm any accuracy the original numbers had: The computer output looks great, but has no meaning.

Studying how errors grow and how to deal with them is usually part of a class in numerical analysis, and Math 443 (Applied Linear Algebra) sometimes includes part of how that works for the kinds of calculations we have been doing. Another aspect of numerical analysis, also covered

in CS classes in computational complexity to some extent, is how to deal with the fact that a large matrix and very large numbers of calculations can be beyond the resources of your computing equipment to do in the simple ways we consider in theory. A process we view as solving the problem may work fine for problems up to some size, but beyond that require either more storage space or more computation time than you can afford. (“Afford”: If you want to deal with a one-million by one-million matrix, 10^{12} numbers each taking several bytes to represent in any given format, there may be no computer hardware that lets you just store them away in the way we have been talking about! Several terabytes of RAM is not just expensive.) I am not going to go into those things here. But I do want you to realize that arithmetic that seems perfectly accurate when we talk about it in a math class, when we refer to it in proofs such as the Gram-Schmidt process, may be very hard to make accurate (or even to carry out at all) in reality. This is relevant to the proof above in the following particular way. I have followed the proof in the textbook, first getting an intermediate basis that was orthogonal but not yet normalized, and then making each vector have norm $\|\vec{v}_i\| = 1$ by dividing by whatever $\|\vec{v}_i\| = \sqrt{(\vec{v}_i, \vec{v}_i)}$ was before normalizing. If you were doing it by hand, it would actually be a little easier to normalize each \vec{v}_i as soon as it is produced, orthogonal to all the preceding ones. If it were, right away, made to have $\|\vec{v}_i\| = 1$, then we could skip dividing by (\vec{v}_i, \vec{v}_i) in finding the later vectors \vec{v}_j , since that would be 1. I.e., do the division by the norm once and not have to do it again each time we use that vector. But depending on what assumptions you make about the size of the various vectors, it may be better to postpone as many divisions as possible as long as possible: The example above calculating w hints at how division can amplify the errors that have crept in, making more error in the inputs to subsequent calculations. So the way we did it in the proof might be better from the standpoint of error growth.

This is really a complex subject and I don’t pretend to have done much with it here, but I did want to call your attention to an aspect of this mathematics that does not appear in our “pure” view where every calculation is done perfectly.