Notes on the Gram-Schmidt Process
MENU, Winter 2013

I'm not too happy with the way in which the book presents the Gram-Schmidt process, and wanted to provide some clarifications as well as an alternate approach. First, recall the goal of the Gram-Schmidt process:

Goal. *Given a collection of linearly independent vectors $\vec{v}_1, \ldots, \vec{v}_k$, produce a collection of orthonormal vectors $\vec{u}_1, \ldots, \vec{u}_k$ with the same span as the original set of vectors.*

Here is (essentially) the way the book describes the process. Start by setting $\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}$. Recall that dividing a vector by its length always produces a unit vector, so $\vec{u}_1$ has length 1 and points in the same direction as $\vec{v}_1$. Then we compute

$$\vec{b}_2 = \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1$$

and set $\vec{u}_2 = \frac{\vec{b}_2}{\|\vec{b}_2\|}$. After the first step above we have produced a vector $\vec{b}_2$ which is orthogonal to $\vec{u}_1$ and the second step (dividing by its length) is to get the result to have length 1. At the next step we would first form

$$\vec{b}_3 = \vec{v}_3 - (\vec{v}_3 \cdot \vec{u}_2)\vec{u}_2 - (\vec{v}_3 \cdot \vec{u}_1)\vec{u}_1$$

and then set $\vec{u}_3 = \frac{\vec{b}_3}{\|\vec{b}_3\|}$. Again, the pattern is that the $\vec{b}_3$ constructed above is orthogonal to $\vec{u}_1$ and $\vec{u}_2$ and then we simply divide by its length to get something of length one. And so on: at the ith step we take $\vec{v}_i$ and subtract off its projections onto all the previous $\vec{u}_j$'s constructed thus far, and then divide the result by its length. Note that at each step we have

$$\operatorname{span}(\vec{u}_1, \ldots, \vec{u}_t) = \operatorname{span}(\vec{v}_1, \ldots, \vec{v}_t),$$

or in other words, at each step the t vectors we have constructed thus far have the same span as the first t vectors among the v_i.

The trouble with this construction is that, since at each step we divide a vector by its length in order to get something of length 1, square roots tend to show up right away and working with these can get kind of messy. Here is an example:

Example 1. Let us find an orthonormal basis for the subspace V of $\mathbb{R}^4$ spanned by the following vectors:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Call these vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ respectively. First we compute:

$$\|\vec{v}_1\| = \sqrt{4} = 2$$

and set

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}.$$

Next we compute

$$\begin{aligned}\vec{b}_2 &= \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1) \cdot \vec{u}_1 \\ &= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{2} \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix} \\ &= \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix},\end{aligned}$$

and divide this by its length in order to get a vector of length 1:

$$\vec{u}_2 = \frac{\vec{b}_2}{\|\vec{b}_2\|} = \frac{1}{\sqrt{12/16}} \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix} = \begin{pmatrix} -3/2\sqrt{3} \\ 1/2\sqrt{3} \\ 1/2\sqrt{3} \\ 1/2\sqrt{3} \end{pmatrix}.$$

As you can see, this is already kind of ugly (even after simplifying a bit), and we're not done yet! Next we would have to compute

$$\vec{b}_3 = \vec{v}_3 - (\vec{v}_3 \cdot \vec{u}_2)\vec{u}_2 - (\vec{v}_3 \cdot \vec{u}_1)\vec{u}_1,$$

giving an ever bigger mess of square roots and fractions.

However, there is a way around this where essentially we perform the Gram-Schmidt process only we don't divide by the length of a vector *until the end* of the entire process. The upshot is that this avoids nasty square roots showing up all over the place, and can make computations easier to follow. So, in this setup we start by setting

$$\vec{b}_1 = \vec{v}_1$$

and leave it as it is, without worrying about the fact (for now) that $\vec{b}_1$ may not have length 1. Next we compute:

$$\vec{b}_2 = \vec{v}_2 - \left(\frac{\vec{v}_2 \cdot \vec{b}_1}{\vec{b}_1 \cdot \vec{b}_1} \right) \vec{b}_1$$

and again leave its length as it is. Notice that we have done here is precisely the same as what we did in the book's description of the Gram-Schmidt process: we have subtracted from $\vec{v}_2$ its projection onto $\vec{b}_1$. The only difference here is that, since $\vec{b}_1$ is not necessarily of length 1, the formula for this projection is

$$\left(\frac{\vec{v}_2 \cdot \vec{b}_1}{\vec{b}_1 \cdot \vec{b}_1} \right) \vec{b}_1$$

instead of the simpler $(\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1$ we had before. At the next step we compute:

$$\vec{b}_3 = \vec{v}_3 - \left(\frac{\vec{v}_3 \cdot \vec{b}_1}{\vec{b}_1 \cdot \vec{b}_1} \right) \vec{b}_1 - \left(\frac{\vec{v}_3 \cdot \vec{b}_2}{\vec{b}_2 \cdot \vec{b}_2} \right) \vec{b}_2,$$

and so on for any additional steps. Now, the vectors $\vec{b}_1, \ldots, \vec{b}_k$ we end up with will no longer necessarily have length 1, but they are still all orthogonal to each other and still have the property that

$$\operatorname{span}(\vec{b}_1, \ldots, \vec{b}_t) = \operatorname{span}(\vec{v}_1, \ldots, \vec{v}_t) \text{ for all } t.$$

Finally, dividing each b_i by its length gives the orthonormal vectors we want:

$$\vec{u}_1 = \frac{\vec{b}_1}{\|\vec{b}_1\|}, \ldots, \vec{u}_k = \frac{\vec{b}_k}{\|\vec{b}_k\|}.$$

So, the point is that instead of after each step producing orthonormal vectors, we first produce vectors which are simply *orthogonal* and only at the end do we ensure they all have length 1. The benefit is that we avoid having to deal with messy square roots until the end. Let's return to the previous example:

Example 2. Take the same vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ as before. Here is what this modified Gram-Schmidt process looks like. First we set

$$\vec{b}_1 = \vec{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Next we compute:

$$\begin{aligned} \vec{b}_2 &= \vec{v}_2 - \left(\frac{\vec{v}_2 \cdot \vec{b}_1}{\vec{b}_1 \cdot \vec{b}_1} \right) \vec{b}_1 \\ &= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}. \end{aligned}$$

Now, as a further simplification, recall that the point of this step is to produce a vector orthogonal to $\vec{b}_1$. The $\vec{b}_2$ thus constructed has this property, but so does *any* multiple of it. Is there a multiple we can use as $\vec{b}_2$ instead to make further computations simpler? Yes! Multiply $\vec{b}_2$ by a scalar to clear all fractions and use that as $\vec{b}_2$ instead:

$$\vec{b}_2 = \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

The point is that not only have we avoided dealing with square roots, but using this idea we can avoid dealing with fractions until absolutely necessary. Since at the end we divide the $\vec{b}$'s by their lengths, replacing one of the $\vec{b}$'s by a multiple of it at some step of the Gram-Schmidt process will not change the final result.

Continuing on, using the new $\vec{b}_2$ above, the next step in the modified Gram-Schmidt process gives:

$$\begin{aligned}
\vec{b}_3 &= \vec{v}_3 - \left(\frac{\vec{v}_3 \cdot \vec{b}_1}{\vec{b}_1 \cdot \vec{b}_1}\right) \vec{b}_1 - \left(\frac{\vec{v}_3 \cdot \vec{b}_2}{\vec{b}_2 \cdot \vec{b}_2}\right) \vec{b}_2 \\
&= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{12} \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{pmatrix}.
\end{aligned}$$

Since we only had three vectors to begin with, we are done with this part of the Gram-Schmidt process, but note that if we had a fourth vector $\vec{v}_4$ to work with, at this point we could replace $\vec{b}_3$ by $3\vec{b}_3$ to avoid dealing with fractions just like we replaced $\vec{b}_2$ earlier.

Finally we divide each of $\vec{b}_1, \vec{b}_2, \vec{b}_3$ by their lengths to give

$$\begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}, \frac{2}{\sqrt{3}} \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}, \sqrt{\frac{3}{2}} \begin{pmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

as the orthonormal vectors we want.

Again, the point of all this is to put off having to deal with any square roots until the very end and to avoid dealing with fractions until absolutely necessary. Feel free to use this modified Gram-Schmidt process if you think it's somewhat simpler to carry out than the book's version.