Notes on the Gramm-Schmidt Procedure for Constructing Orthonormal Bases

by

Eric Carlen

Suppose I’m given two vectors, say

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix},$$

and I want to find an orthonormal basis for the subspace $V$ that they span. Here is a standard procedure, called “Gram–Schmidt orthogonalization” for doing this.

First, define the first element, $u_1$, of our basis to be the normalization of $v_1$. That is, we rescale $v_1$ to make it a unit vector:

$$u_1 = \frac{1}{|v_1|}v_1.$$

Orthogonality, like a tango, takes two. So at this stage we don’t need to worry about orthogonality, just normality. But to define $u_2$, it is a different story. We can’t just normalize $v_2$, because the result might not be orthogonal to $u_1$.

So here is what we will do: Let’s construct a vector $w_2$ that is orthogonal to $u_1$, and let’s build it out of $v_2$ and $u_1$. Let’s define

$$w_2 = v_2 - au_1$$

and let’s choose $a$ to make $w_2$ orthogonal to $u_1$.

To get an equation for what $a$ should be, we use the fact that two vectors are orthogonal if and only if their dot product is zero. So we must choose $a$ to make

$$0 = w_2 \cdot u_1 = (v_2 - au_1) \cdot u_1 = v_2 \cdot u_1 - a|u_1|^2 = v_2 \cdot u_1 - a.$$

Evidently, the choice $a = v_2 \cdot u_1$ makes $w_2$ orthogonal to $u_1$, and so

$$w_2 = v_2 - (v_2 \cdot u_1)u_1.$$

Now, provided $|w_2| \neq 0$, we can normalize $w_2$ and make it a unit vector. Moreover, in rescaling it, we won’t change the fact that it is orthogonal to $u_1$. So we define

$$u_2 = \frac{1}{|w_2|}w_2.$$

Now what about that “provided”? If $w_2 = 0$, then

$$v_2 = (v_2 \cdot u_1)u_1 = \frac{v_2 \cdot u_1}{|v_1|}v_1.$$
This could only happen if $v_2$ were a multiple of $v_1$, but clearly in our example it isn’t, so nothing will go wrong there.

Now let’s work it out. Clearly $|v_1| = \sqrt{2}$ so that

$$u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$ 

Next, we compute that

$$v_2 \cdot u_1 = \frac{3}{\sqrt{2}}$$

and hence

$$w_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}.$$ 

Finally,

$$u_2 = \frac{1}{|w_2|}w_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

and our orthonormal basis is:

$$u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad u_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}. $$

Now let’s move on to working with more than two vectors. Given a collection

$$\{v_1, v_2, \ldots, v_k\}$$

of non-zero $n$ dimensional vectors we define $u_1$ and $u_2$ just as above.

$$u_1 = \frac{1}{|v_1|}v_1$$

and

$$u_2 = \frac{1}{|w_2|}w_2 \quad \text{where} \quad w_2 = v_2 - (v_2 \cdot u_1)u_1.$$ 

If $|w_2| = 0$, then you can’t divide by it. In case this happens, just throw $v_2$ out of the list, and relabel the subscripts so that they are consecutive; i.e., $v_3$ becomes the new $v_2$, etc. Now try again. Keep going until you get a $u_2$, or else run out of vectors.

If we run out of vectors, we are done, so suppose we didn’t run out of vectors. After relabeling the ones we’ve still got, the next one is $v_3$. Define

$$w_3 = v_3 - (v_3 \cdot u_1)u_1 - (v_3 \cdot u_2)u_2.$$
You can easily check that since
\[ u_1 \cdot u_1 = u_2 \cdot u_2 = 1 \quad \text{and} \quad u_1 \cdot u_2 = 0 \]
that
\[ w_3 \cdot u_1 = w_3 \cdot u_2 = 0 \]
so that \( w_3 \) is orthogonal to \( u_1 \) and \( u_2 \). So as long as \(|w_3| > 0\), we can divide by it and define:
\[ u_3 = \frac{1}{|w_3|} w_3 \quad \text{where} \quad w_3 = v_3 - (v_3 \cdot u_1)u_1 - (v_3 \cdot u_2)u_2. \]

By now you see the pattern: Assuming that either no vectors have been thrown out, or that they have been relabeled to keep the subscripts consecutive after each time one is thrown out,
\[ u_j = \frac{1}{|w_j|} w_j \quad \text{where} \quad w_j = v_j - \sum_{i=1}^{j-1} (v_j \cdot u_i)u_i. \]

**Theorem 1 (Gram-Schmidt Procedure)** The Gram-Schmidt process always produces an orthonormal basis for the span of
\[ \{v_1, v_2, \ldots, v_k\}. \]

The main point in the proof of this theorem will be to show that \( \{u_1, u_2, \ldots, u_\ell\} \) spans the span of \( \{v_1, v_2, \ldots, v_k\} \), so we now establish a simple test for this in the case \( \ell = k \) which is all we will need. Suppose \( \ell = k \) and let \( A \) be the \( n \times k \) matrix whose \( j \)-th column is \( v_j \), and let \( B \) be the \( n \times k \) matrix whose \( j \)-th column is \( u_j \). Then the span of the \( \{v_1, v_2, \ldots, v_k\} \) is the column space of \( A \), and the span of \( \{u_1, u_2, \ldots, u_k\} \) is the column space of \( B \). So what we are after is a test for when two \( n \times k \) matrices have the same column space.

**Theorem 2 (Equality of Column Spaces)** Let \( A \) and \( B \) be two \( n \times k \) matrices. If there is an invertible \( k \times k \) matrix \( C \) so that
\[ A = BC, \]
then the column spaces of \( A \) and \( B \) are the same.

**Proof:** First suppose that \( A = BC \) and that \( C \) is invertible. A \( n \)-dimensional vector \( w \) belongs to the column space of \( A \) if and only if there is a \( k \)-dimensional vector \( x \) so that \( w = Ax \). But then since \( A = BC \),
\[ w = Ax = B(Cx) \]

3
and so there is a vector \( y \) with \( w = By \), namely \( y = Cx \). So \( w \) belongs to the column space of \( B \). Since \( w \) was an arbitrary element of the column space of \( A \), we have that every element of the column space of \( A \) is in the column space of \( B \). That is, the column space of \( A \) is contained in the column space of \( B \).

Now write \( B = AC^{-1} \) and repeat the argument with the roles of \( A \) and \( B \) switched. This yields that the column space of \( B \) is contained in the column space of \( A \). If two spaces are contained in each other, then they are the same, and so we have proved that the column spaces are equal.

**Proof of Theorem 1:** Let \( \{u_1, u_2, \ldots, u_k\} \) be the orthonormal vectors produced by the Gram-Schmidt procedure. They are orthonormal by the construction, so they are certainly independent. The question is: “Do they span the same space as the original vectors \( \{v_1, v_2, \ldots, v_k\} \)?” The theorem says “yes”, so that’s what we have to show.

Before we begin, notice that \( \ell \leq k \), and \( \ell < k \) if and only if one or more of the original vectors was thrown out to avoid dividing by zero.

First let’s consider the case in which nothing gets thrown out; i.e., \( \ell = k \). Proceeding under this assumption, we observe that

\[
    v_j = w_j + \sum_{1=1}^{j-1} (v_j \cdot u_i)u_i
\]

\[
    = |w_j|u_j + \sum_{1=1}^{j-1} (v_j \cdot u_i)u_i
\]

and so

\[
    v_j = \sum_{1=1}^{j} R_{ij}u_i
\]

where

\[
    R_{jj} = |w_j| \quad \text{and} \quad R_{ij} = v_j \cdot u_i \quad \text{for} \quad i < j
\]

Now let’s define

\[
    R_{ij} = 0 \quad \text{for} \quad i > j
\]

Then the \( R_{ij} \) are the entries of a \( k \) by \( k \) matrix \( R \) and

\[
    v_j = \sum_{1=1}^{k} R_{ij}u_i.
\]

This means that if \( A \) is the \( n \) by \( k \) matrix whose \( j \)th column is \( v_j \), and if \( Q \) is the \( n \) by \( k \) matrix whose \( j \)th column is \( u_j \), then

\[
    A = QR.
\]
Now notice that $R$ is an upper–triangular matrix: all entries below the main diagonal are zero. So it is already partially row-reduced, and it is invertible if and only if none of its diagonal entries are zero. But by assumption,

$$R_{ii} = |w_i| > 0$$

so $R$ is invertible. Now Theorem 2 says that the column spaces of $A$ and $Q$ are the same. But these are justs the spans of $\{v_1, v_2, \ldots, v_k\}$ and $\{u_1, u_2, \ldots, u_k\}$. This completes the proof in the case $\ell = k$.

Now suppose $\ell < k$, which means that some of the $v_j$ get thrown out. We throw $v_j$ out only if

$$v_j - \sum_{i=1}^{j-1} (v_j \cdot u_i)u_i = 0 .$$

But this means that $v_j$ is in the span of $\{v_1, v_2, \ldots, v_{j-1}\}$, and so we have the same span after throwing out $v_j$ as we did before. So we may as well remove every vector $v_j$ that can be written as a linear combination of the $\{v_1, v_2, \ldots, v_{j-1}\}$ at the outset. The remaining vectors have the same span as the original set. Now apply the Gram–Schmidt procedure to the reduced set. Now nothing gets thrown out, and by the first part of the proof, the result is an orthonormal set of vectors that has the same span as the reduced set, and finally the original set. So in either case, the span is the same, and the theorem is proved.

As a final example, let’s use the Gram–Schmidt procedure to compute an orthonormal basis for the column space of the 4 by 3 matrix $A$ where

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \\ 2 & 3 & 0 \end{bmatrix} .$$

Let $v_1$, $bv_2$, and $bv_3$ be the columns of $A$, ordered left to right as usual. Then $|v_1| = \sqrt{6}$ so that

$$u_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} .$$

Next, we compute that

$$v_2 \cdot u_1 = \frac{9}{\sqrt{6}} = \frac{3\sqrt{3}}{\sqrt{2}}$$

and hence

$$w_2 = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \end{bmatrix} - \frac{3\sqrt{3}}{\sqrt{2}} \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 4 \\ 0 \\ -1 \end{bmatrix}$$
and so
\[
\mathbf{u}_2 = \frac{1}{\|\mathbf{w}_2\|} \mathbf{w}_2 = \frac{1}{3\sqrt{2}} \begin{bmatrix} 1 \\ 4 \\ 0 \\ -1 \end{bmatrix}
\]

Next, we compute that
\[
\mathbf{v}_3 \cdot \mathbf{u}_1 = \frac{\sqrt{2}}{\sqrt{3}}
\]
and that
\[
\mathbf{v}_3 \cdot \mathbf{u}_2 = \frac{1}{3\sqrt{2}} 4 = \frac{2\sqrt{2}}{3}.
\]

Therefore,
\[
\mathbf{w}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{\sqrt{2}}{\sqrt{3}\sqrt{6}} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{2\sqrt{2}}{3\sqrt{2}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} \\ 0 \\ -6 \end{bmatrix}
\]

Finally,
\[
\mathbf{u}_3 = \frac{1}{\sqrt{117}} \begin{bmatrix} 4 \\ 1 \\ -6 \\ 8 \end{bmatrix}
\]
is the third element of our orthonormal basis. Evidently, the dimension of the column space of \(A\) is 3.

**Problems**

(1) Find an orthonormal basis for the column space of the matrix
\[
A = \begin{bmatrix} 1 & 2 & 4 & 17 \\ 0 & 2 & 2 & 0 \\ 2 & 3 & 7 & 1 \\ 1 & 1 & 3 & 0 \end{bmatrix}.
\]
what is the dimension of this column space?
(2) Find an orthonormal basis for the column space transpose $A^t$ of the matrix $A$ from problem (1). From this, you can easily deduce an orthonormal basis for the row space of the original matrix $A$. What is it, and what is the dimension of the row space?

(3) Let $A$ be the matrix

$$
A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 2 \\ 2 & 0 & 1 \end{bmatrix}.
$$

Find a 3 by 3 upper triangular matrix $R$ and a 3 by 3 matrix $Q$ whose columns are orthonormal such that

$$
A = QR
$$

as in the proof of Theorem 1. Check your result by computing the product $QR$. 