Review. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthogonal, the orthogonal projection of $\boldsymbol{w}$ onto $\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is

$$\hat{\boldsymbol{w}} = \frac{\boldsymbol{w}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1 + \ldots + \frac{\boldsymbol{w}\cdot\boldsymbol{v}_n}{\boldsymbol{v}_n\cdot\boldsymbol{v}_n}\boldsymbol{v}_n.$$

Example 64.

(a) Project $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ onto $W = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}\right\}$.

(b) Express $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ in terms of the basis $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$.

Solution.

(a) The projection is $\frac{8}{6}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \frac{4}{5}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$. (Each coefficient is obtained as the quotient of two dot products.)

(b) $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \frac{8}{6}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \frac{4}{5}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + \frac{5}{30}\begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$

Gram–Schmidt

(Gram–Schmidt orthogonalization)
Given a basis $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots$ for W, we produce an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ for W as follows:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \begin{pmatrix}\text{projection of} \\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1\end{pmatrix}$
- $\boldsymbol{q}_3 = \boldsymbol{w}_3 - \begin{pmatrix}\text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1\end{pmatrix} - \begin{pmatrix}\text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2\end{pmatrix}$
- $\boldsymbol{q}_4 = \ldots$

Note. Since $\boldsymbol{q}_1, \boldsymbol{q}_2$ are orthogonal, $\begin{pmatrix}\text{projection of} \\ \boldsymbol{w}_3 \text{ onto span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\}\end{pmatrix} = \begin{pmatrix}\text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1\end{pmatrix} + \begin{pmatrix}\text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2\end{pmatrix}$.

Important comment. When working numerically on a computer it actually saves time to compute an orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ by the same approach but always normalizing each $\boldsymbol{q}_i$ along the way. The reason this saves time is that now the projections onto $\boldsymbol{q}_i$ only require a single dot product (instead of two). This is called **Gram–Schmidt orthonormalization**. When working by hand, it is usually simpler to wait until the end to normalize (so as to avoid).

Note. When normalizing, the orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ is the unique one (up to $\pm$ signs) with the property that $\operatorname{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_k\} = \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ for all $k = 1, 2, \ldots$.

Armin Straub
straub@southalabama.edu

Example 65. Using Gram–Schmidt, find an orthogonal basis for $W = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$.

Solution. We already have the basis $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{w}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ for W. However, that basis is not orthogonal.

We can construct an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2$ for W as follows:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}$

  Note. $\boldsymbol{q}_2$ is the error of the projection of $\boldsymbol{w}_2$ onto $\boldsymbol{q}_1$. This guarantees that it is orthogonal to $\boldsymbol{q}_1$. On the other hand, since $\boldsymbol{q}_2$ is a combination of $\boldsymbol{w}_2$ and $\boldsymbol{q}_1$, we know that $\boldsymbol{q}_2$ actually is in W.

We have thus found the orthogonal basis $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{2}{3}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ for W (if we like, we can, of course, drop that $\frac{2}{3}$).

Important comment. By normalizing, we get an orthonormal basis for W: $\frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

Practical comment. When implementing Gram–Schmidt on a computer, it is beneficial (slightly less work) to normalize each $\boldsymbol{q}_i$ during the Gram–Schmidt process. This typically introduces square roots, which is why normalizing at the end is usually preferable when working by hand.

Comment. There are, of course, many orthogonal bases $\boldsymbol{q}_1, \boldsymbol{q}_2$ for W. Up to the length of the vectors, ours is the unique one with the property that $\text{span}\{\boldsymbol{q}_1\} = \text{span}\{\boldsymbol{w}_1\}$ and $\text{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\} = \text{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$.

A matrix Q has orthonormal columns $\iff$ $Q^T Q = I$

Why? Let $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ be the columns of Q. By the way matrix multiplication works, the entries of $Q^T Q$ are dot products of these columns:

$$\begin{bmatrix} — & \boldsymbol{q}_1^T & — \\ — & \boldsymbol{q}_2^T & — \\ & \vdots & \end{bmatrix} \begin{bmatrix} | & | & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

Hence, $Q^T Q = I$ if and only if the dot products $\boldsymbol{q}_i^T \boldsymbol{q}_j = 0$ (that is, the columns are orthogonal), for $i \neq j$, and $\boldsymbol{q}_i^T \boldsymbol{q}_i = 1$ (that is, the columns are normalized).

Example 66. $Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$ obtained from Example 65 satisfies $Q^T Q = I$.

The QR decomposition

Just like the LU decomposition encodes the steps of Gaussian elimination, the QR decomposition encodes the steps of Gram–Schmidt.

(QR decomposition) Every $m\times n$ matrix A of rank n can be decomposed as $A=QR$, where

- Q has orthonormal columns, $(m\times n)$
- R is upper triangular and invertible. $(n\times n)$

How to find Q and R?

- Gram–Schmidt orthonormalization on (columns of) A, to get (columns of) Q
- $R=Q^TA$
  Why? If $A=QR$, then $Q^TA=Q^TQR$ which simplifies to $R=Q^TA$ (since $Q^TQ=I$).

The decomposition $A=QR$ is unique if we require the diagonal entries of R to be positive (and this is exactly what happens when applying Gram–Schmidt).

Practical comment. Actually, no extra work is needed for computing R. All of its entries have been computed during Gram–Schmidt.

Variations. We can also arrange things so that Q is an $m\times m$ orthogonal matrix and R a $m\times n$ upper triangular matrix. This is a tiny bit more work (and not required for many applications): we need to complement "our" Q with additional orthonormal columns and add corresponding zero rows to R. For square matrices this makes no difference.

Example 67. Determine the QR decomposition of $A=\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}$.

Solution. The first step is Gram–Schmidt orthonormalization on the columns of A. We then use the resulting orthonormal vectors (they need to be normalized!) as the columns of Q.

We already did Gram–Schmidt in Example 65: from that work, we have $Q=\begin{bmatrix}1/\sqrt{3}&1/\sqrt{6}\\1/\sqrt{3}&-2/\sqrt{6}\\1/\sqrt{3}&1/\sqrt{6}\end{bmatrix}$.

Hence, $R=Q^TA=\begin{bmatrix}1/\sqrt{3}&1/\sqrt{3}&1/\sqrt{3}\\1/\sqrt{6}&-2/\sqrt{6}&1/\sqrt{6}\end{bmatrix}\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}=\begin{bmatrix}3/\sqrt{3}&1/\sqrt{3}\\0&4/\sqrt{6}\end{bmatrix}$.

Comment. As commented earlier, the entries of R have actually all been computed during Gram–Schmidt, so that, if we pay attention, we could immediately write down R (no extra work required). Looking back at Example 65, can you see this?

Check. Indeed, $QR=\begin{bmatrix}1/\sqrt{3}&1/\sqrt{6}\\1/\sqrt{3}&-2/\sqrt{6}\\1/\sqrt{3}&1/\sqrt{6}\end{bmatrix}\begin{bmatrix}3/\sqrt{3}&1/\sqrt{3}\\0&4/\sqrt{6}\end{bmatrix}=\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}$ equals A.