Example 65. Using Gram–Schmidt, find an orthogonal basis for $W = \text{span}\left\{ \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Solution. Let us repeat the previous step so the entire procedure becomes more transparent.

We begin with the (not orthogonal) basis $\boldsymbol{w}_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}$, $\boldsymbol{w}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\boldsymbol{w}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

We then construct an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ as follows:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{3}{9} \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
- $\boldsymbol{q}_3 = \boldsymbol{w}_3 - \begin{pmatrix} \text{projection of } \boldsymbol{w}_3 \\ \text{onto span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\} \end{pmatrix} = \boldsymbol{w}_3 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2 \end{pmatrix}$
  $= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{9} \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix} - \frac{2}{4} \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

  Make sure you see why $\boldsymbol{q}_3$ is orthogonal to both $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$!

  Also note that breaking up the projection onto $\text{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$ into the projections onto $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ is only possible because $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are orthogonal.

Indeed, $\begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ is an orthogonal basis of $\mathbb{R}^3$.

If we prefer, we can normalize to obtain an orthonormal basis: $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$.

It is common and beneficial (slightly less work) to normalize during the Gram–Schmidt process. We do this in Example 66 below.

The following is just the Gram–Schmidt orthogonalization except that we immediately normalize each vector $\boldsymbol{q}_i$.

(Gram–Schmidt orthonormalization)

Given a basis $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots$ for W, produce an orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ for W.

- $\boldsymbol{q}_1 = \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|}$ with $\boldsymbol{b}_1 = \boldsymbol{w}_1$
- $\boldsymbol{q}_2 = \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|}$ with $\boldsymbol{b}_2 = \boldsymbol{w}_2 - (\boldsymbol{w}_2 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1$
- $\boldsymbol{q}_3 = \frac{\boldsymbol{b}_3}{\|\boldsymbol{b}_3\|}$ with $\boldsymbol{b}_3 = \boldsymbol{w}_3 - (\boldsymbol{w}_3 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1 - (\boldsymbol{w}_3 \cdot \boldsymbol{q}_2)\boldsymbol{q}_2$
- $\boldsymbol{q}_4 = \ldots$

Example 66. Find an orthonormal basis for $W = \text{span}\left\{ \begin{bmatrix} 0\\3\\0\\0 \end{bmatrix}, \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} \right\}$.

Solution. Let $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ be the vectors spaning W. We then construct an orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ using Gram–Schmidt orthonormalization as follows:

- $\boldsymbol{b}_1 = \begin{bmatrix} 0\\3\\0\\0 \end{bmatrix}$, so that $\boldsymbol{q}_1 = \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}$.
- $\boldsymbol{b}_2 = \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix} - \left(\begin{bmatrix} 2\\1\\0\\0 \end{bmatrix} \cdot \boldsymbol{q}_1 \right) \boldsymbol{q}_1 = \begin{bmatrix} 2\\0\\0\\0 \end{bmatrix}$, so that $\boldsymbol{q}_2 = \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}$.
- $\boldsymbol{b}_3 = \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} - \left(\begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} \cdot \boldsymbol{q}_1 \right) \boldsymbol{q}_1 - \left(\begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} \cdot \boldsymbol{q}_2 \right) \boldsymbol{q}_2 = \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}$, so that $\boldsymbol{q}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}$.

We have found the orthonormal basis: $\begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}$ (which, of course, matches the previous example).

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 67. $Q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}$ obtained from the previous example satisfies $Q^T Q = I$.

Armin Straub
straub@southalabama.edu

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
- $R = Q^T A$
  Why? If $A = QR$, then $Q^T A = Q^T QR$ which simplifies to $R = Q^T A$ (since $Q^T Q = I$).

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 68. Determine the QR decomposition of $A = \begin{bmatrix} 0 & 2 & 1 \\ 3 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Solution. The first step is Gram–Schmidt orthonormalization on the columns of A. We then use the resulting orthonormal vectors as the columns of Q.

We already did Gram–Schmidt in Example 66: from that work, we have $Q = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}$.

Hence, $R = Q^T A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 2 & 1 \\ 3 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$.

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 66, can you see this?

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

Example 69. (extra) Find the QR decomposition of $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

Solution. (final answer only) $A = QR$ with $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$ and $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$.

Example 70. One practical application of the QR decomposition is solving systems of linear equations.

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} \quad &\Longleftrightarrow \quad QR\boldsymbol{x} = \boldsymbol{b} \qquad \text{(now, multiply with } Q^T \text{ from the left)} \\ &\Longrightarrow \quad R\boldsymbol{x} = Q^T\boldsymbol{b} \end{aligned}$$

The last system is triangular and can be solved by back substitution.

A couple of comments are in order:

- If A is $n \times n$ and invertible, then the "$\Longrightarrow$" is actually a "$\Longleftrightarrow$".
- The equation $R\boldsymbol{x} = Q^T\boldsymbol{b}$ is always consistent! (Recall that R is invertible.)

  Indeed, if A is not $n \times n$ or not invertible, then $R\boldsymbol{x} = Q^T\boldsymbol{b}$ gives the least squares solutions!

  Why? $A^TA\hat{\boldsymbol{x}} = A^T\boldsymbol{b} \quad \Longleftrightarrow \quad \underbrace{(QR)^TQR}_{=R^TQ^TQR}\hat{\boldsymbol{x}} = (QR)^T\boldsymbol{b} \quad \Longleftrightarrow \quad R^TR\hat{\boldsymbol{x}} = R^TQ^T\boldsymbol{b} \quad \Longleftrightarrow \quad R\hat{\boldsymbol{x}} = Q^T\boldsymbol{b}$

  [For the last step we need that R is invertible, which is always the case when A is $m \times n$ of rank n.]
- So, how does the QR way of solving linear systems compare to our beloved Gaussian elimination (LU)?

  It turns out that QR is a little slower than LU but makes up for it in "numerical stability".

  What does that mean? When computing numerically, we use floating point arithmetic and approximate each number by an expression of the form $0.1234 \cdot 10^{-16}$. A certain (fixed) number of bits is used to store the part 0.1234 (here, 4 decimal places of accuracy) as well as the exponent -16.

  Now, here is something terrible that can happen in numerical computations: mathematically, the quantities x and $(x+1)-1$ are exactly the same. However, numerically, they might not. Take, for instance, $0.1234 \cdot 10^{-6}$. Then, to an accuracy of 4 decimal places, $x + 1 = 0.1000 \cdot 10^1$, so that $(x+1) - 1 = 0.0000$. But $x \neq 0$. We completely lost all the information about x.

  To be numerically stable, an algorithm must avoid issues like that.

$\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x} = \boldsymbol{b}$
$\Longleftrightarrow \quad R\hat{\boldsymbol{x}} = Q^T\boldsymbol{b}$ (where $A = QR$)

Armin Straub
straub@southalabama.edu