Gram-Schmidt and QR

The Gram-Schmidt process (GSp) takes a sequence $a_1, a_2, \ldots, a_n$ of linearly independent vectors and gives a sequence $q_1, q_2, \ldots, q_n$ of orthonormal vectors which satisfy

$$\mathrm{Span}(q_1, q_2, \ldots, q_k) = \mathrm{Span}(a_1, a_2, \ldots, a_k), \quad k = 1, 2, \ldots, n. \tag{1}$$

Everything about the GSp is here in (1). You can even see how it works: Suppose you have $q_1, q_2, \ldots, q_{k-1}$, and you want q_k. Since we want $a_k \in \mathrm{Span}(q_1, q_2, \ldots, q_k)$, we write

$$a_k = r_{kk} q_k + \sum_{j=1}^{k-1} r_{jk} q_j. \tag{2}$$

Premultiplying by q_j^t (and noting orthogonality) gives

$$r_{jk} = q_j^t a_k, \quad j = 1, 2, \ldots, k-1, \tag{3}$$

and now that the r_{ij} are known we can use (2) to define

$$w \equiv r_{kk} q_k = a_k - \sum_{j=1}^{k-1} r_{jk} q_j. \tag{4}$$

This gives the direction of q_k, and r_{kk} is chosen (usually positive) so that q_k has unit length:

$$r_{kk} = \|w\|_2, \tag{5}$$

and

$$q_k = w / r_{kk}. \tag{6}$$

The GSp is simply (3), (4), (5), and (6) for $k = 1, 2, \ldots, n$.

The geometry of the k^{th} step of GSp is simple:

> For each $j = 1, 2, \ldots, k-1$, (3) and (4) subtracts from a_k its projection onto q_j. The resulting vector is then orthogonal to q_j. The final vector, w, is then orthogonal to $q_1, q_2, \ldots, q_{k-1}$. Interpret r_{jk} as the (signed) length of the projection of a_k onto q_j.
>
> Steps (5) and (6) simply normalize the new q_k.

Notice that the only time that something can go wrong here is if $w = 0$; that is, if after subtracting the projections of a_k onto the previous q_j, we end up with nothing. But that means that a_k is a linear combination of the previous q_j, and thus that the a_k are linearly dependent.

If we have coordinates, let $A = [a_1, a_2, \ldots, a_n]$, let $Q = [q_1, q_2, \ldots, q_n]$ and let R be the $n \times n$ upper triangular matrix defined by (3) and (5). Then the $m \times n$ matrix Q has orthonormal columns, and $A = QR$. This is called a (thin) QR factorization of A. The columns of Q form an orthonormal basis for ColSp(A).

Notice also that (if the columns of A are linearly independent) the only freedom we have above is the sign of r_{kk} (or its phase if it were complex). Thus, except for the sign of the r_{kk} and (therefore) the *sense* of the q_k, the QR factorization is unique. We say that if A has full column rank, the QR factorization is essentially unique. While there are other ways to compute it, the GSp essentially defines the (thin) QR factorization.