Addendum to Section 8.3.3: QR and Least-Squares

In the text the use of a QR factorization for solving least-squares problems is based on the development of QR in Chapter 4. In what follows the factorization is developed from scratch, using an approach tailored to how it is used for least-squares.

If $\mathbf{A}$ is $n \times m$, then it has a QR factorization of the form

$$\mathbf{A} = \mathbf{QR}, \tag{1}$$

where $\mathbf{Q}$ is an $n \times n$ orthogonal matrix, and $\mathbf{R}$ is an $n \times m$ upper triangular matrix. Although all matrices have a QR factorization, it is assumed in what follows that the columns of $\mathbf{A}$ are independent.

The key step is to replace the columns of $\mathbf{A}$ with orthonormal equivalents, and how this can be done is illustrated in the following example.

Example

Suppose that $\mathbf{A}$ has 3 columns and n rows. So, it is possible to write $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{pmatrix}$, where $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are the 3 column vectors from $\mathbf{A}$. With this, given any $\mathbf{v} = (v_1, v_2, v_3)^T$,

$$\mathbf{Av} = v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3\mathbf{a}_3.$$

This shows that $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ is a basis for the range of $\mathbf{A}$.

Step 1: Construct an Orthonormal Basis

We are going to replace $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ with an orthonormal basis $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ using Gram-Schmidt. The steps are as follows:

I) Find orthogonal replacements $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$: This can be done by taking

$$\begin{aligned}
\mathbf{e}_1 &= \mathbf{a}_1, \\
\mathbf{e}_2 &= \mathbf{a}_2 - \alpha_2\mathbf{a}_1, \\
\mathbf{e}_3 &= \mathbf{a}_3 - \alpha_3\mathbf{a}_1 - \beta_3\mathbf{a}_2,
\end{aligned}$$

where α_2, α_3, and β_3 are used to satisfy the orthogonality requirements. For example, to have $\mathbf{e}_2 \cdot \mathbf{e}_1 = 0$ requires that $\alpha_2 = \mathbf{a}_2 \cdot \mathbf{a}_1/\mathbf{a}_1 \cdot \mathbf{a}_1$. Similarly, α_3 and β_3 are determined from the requirements that $\mathbf{e}_3 \cdot \mathbf{e}_1 = 0$ and $\mathbf{e}_3 \cdot \mathbf{e}_2 = 0$ (the exact formulas for α_3 and β_3 are not needed in what follows).

II) Normalize to obtain $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$: Let

$$\begin{aligned}
\mathbf{q}_1 &= \gamma_1\mathbf{e}_1, \\
\mathbf{q}_2 &= \gamma_2\mathbf{e}_2, \\
\mathbf{q}_3 &= \gamma_3\mathbf{e}_3,
\end{aligned}$$

where $\gamma_1 = 1/||\mathbf{e}_1||_2$, $\gamma_2 = 1/||\mathbf{e}_2||_2$, and $\gamma_3 = 1/||\mathbf{e}_3||_2$.

Writing out the resulting expressions we have

$$\begin{aligned}
\mathbf{q}_1 &= \gamma_1 \mathbf{a}_1, \\
\mathbf{q}_2 &= \gamma_2(\mathbf{a}_2 - \alpha_2 \mathbf{a}_1), \\
\mathbf{q}_3 &= \gamma_3(\mathbf{a}_3 - \alpha_3 \mathbf{a}_1 - \beta_3 \mathbf{a}_2).
\end{aligned}$$

Solving for $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ we get

$$\mathbf{a}_1 = r_{11}\mathbf{q}_1, \tag{2}$$

$$\mathbf{a}_2 = r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2, \tag{3}$$

$$\mathbf{a}_3 = r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3, \tag{4}$$

where $r_{11} = 1/\gamma_1$, $r_{12} = \alpha_2/\gamma_1$, etc.

Step 2: Factor A

Substituting (2)-(4) into $\mathbf{A}$, we get

$$\begin{aligned}
\mathbf{A} &= \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{pmatrix} \\
&= \begin{pmatrix} r_{11}\mathbf{q}_1 & r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 & r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3 \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{pmatrix} \\
&= \mathbf{G}\mathbf{R}_m \,, \qquad (5)
\end{aligned}$$

where

$$\mathbf{G} = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{pmatrix}, \quad \text{and} \quad \mathbf{R}_m = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{pmatrix}. \tag{6}$$

The expression in (5) is known as a thin QR factorization of $\mathbf{A}$. How it differs from a (full) QR factorization is explain later.

It is evident from (2)-(4) that the upper triangular form of $\mathbf{R}_m$ is due to how Gram-Schmidt constructs $\mathbf{q}_1$, $\mathbf{q}_2$, and $\mathbf{q}_3$. If you were to randomly pick an orthonormal basis for the range it is very likely that $\mathbf{a}_1$ would also depend $\mathbf{q}_2$ and $\mathbf{q}_3$, and $\mathbf{a}_2$ would also depend $\mathbf{q}_3$. In such a case, $\mathbf{R}_m$ in (6) would not be upper triangular. ■

Once you know how a thin QR factorization of $\mathbf{A}$ is determined, you can show that there is an easier way to find $\mathbf{R}_m$. This comes from the observation that $\mathbf{G}^T\mathbf{G} = \mathbf{I}$. So, multiplying $\mathbf{A} = \mathbf{G}\mathbf{R}_m$ by $\mathbf{G}^T$ we get $\mathbf{R}_m = \mathbf{G}^T\mathbf{A}$. This formula can be used to find $\mathbf{R}_m$. Also, in what follows we write $\mathbf{G} = \mathrm{GS}(\mathbf{A})$. In words, this states that $\mathbf{G}$ is determined by applying the Gran-Schmidt procedure to the columns of $\mathbf{A}$.

Summary: Thin QR Factorization

To summarize, assuming the columns of $\mathbf{A}$ are independent, then the factorization $\mathbf{A} = \mathbf{G}\mathbf{R}_m$ is obtained as follows:

$$\mathbf{G} = \mathrm{GS}(\mathbf{A}),$$

$$\mathbf{R}_m = \mathbf{G}^T\mathbf{A}.$$

This is called a *thin*, or *reduced*, *QR factorization.* With it, the solution of the linear least squares problem, given in (8.15) of the text, is found by solving $\mathbf{R}_m\mathbf{v} = \mathbf{y}_m$, where $\mathbf{y}_m = \mathbf{G}^T\mathbf{y}$. An illustration of this approach is given in Example 8.4 of the text.

A Full QR Factorization

The matrix $\mathbf{G}$ has orthonormal columns but it is not orthogonal unless $n = m$. It is not difficult to modify it obtain a QR factorization as given in (1). To explain how, suppose in the above example that $n = 5$. Given we have 3 orthonormal vectors $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ in $\mathbb{R}^5$, it is possible to find two more vectors $\mathbf{q}_4$, $\mathbf{q}_5$ so that $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, $\mathbf{q}_4, \mathbf{q}_5$ is an orthonormal basis of $\mathbb{R}^5$. With this,

$$\mathbf{Q} = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 & \mathbf{q}_4 & \mathbf{q}_5 \end{pmatrix}$$

is our sought after orthogonal matrix. The modification to turn $\mathbf{R}_m$ into $\mathbf{R}$ is to add the same number of rows to $\mathbf{R}_m$ as we added columns to $\mathbf{G}$. In this case the new rows are all zeros. So, for the above example we get

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

There are various ways to determine $\mathbf{q}_4$ and $\mathbf{q}_5$. However, as illustrated in Example 8.4 in the text, for least-squares all that is needed is a thin QR factorization. So, we won't work out any examples doing this.

When determining either a thin or full QR by hand, the procedure outlined above is just fine. However, when using a computer (i.e., floating point arithmetic), there are concerns about Gram-Schmidt and these are discussed in Section 4.4.2. For this reason, QR factorizations are often computed using the Householder approach and this is explained in Section 8.4.