Lecture 4, Gram-Schmidt, QR solution to least squares

Michael S. Floater

September 19, 2018

These notes are based on Sections 2.6 and 2.9 of the book.

1 The Gram-Schmidt Algorithm

There is no unique way to construct a QR factorization. The simplest, conceptually, is the Gram-Schmidt algorithm. We continue to assume that the m columns of A are linearly independent. The initial task is to find orthonormal $\mathbf{q}_1, \ldots, \mathbf{q}_m$ and coefficients $R_{i,k}$ such that

$$\mathbf{a}_k = \sum_{i=1}^{k} R_{i,k} \mathbf{q}_i, \qquad k = 1, \ldots, m, \tag{1}$$

First, the case $k = 1$ of (1) is

$$\mathbf{a}_1 = R_{1,1} \mathbf{q}_1.$$

Since $\mathbf{a}_1 \neq 0$, we can solve this by letting $\mathbf{q}_1 = \mathbf{a}_1 / \|\mathbf{a}_1\|$ and $R_{1,1} = \|\mathbf{a}_1\|$.

Next, the case $k = 2$ of (1) is

$$\mathbf{a}_2 = R_{1,2} \mathbf{q}_1 + R_{2,2} \mathbf{q}_2. \tag{2}$$

To solve this, we form the vector

$$\mathbf{w} = \mathbf{a}_2 - \langle \mathbf{q}_1, \mathbf{a}_2 \rangle \mathbf{q}_1.$$

It is by construction orthogonal to $\mathbf{q}_1$. Since $\mathbf{a}_1, \mathbf{a}_2$ are linearly independent, $\mathbf{w} \neq 0$. Then we set $\mathbf{q}_2 = \mathbf{w}/\|\mathbf{w}\|$. Then, $\mathbf{q}_1$ and $\mathbf{q}_2$ are orthonormal. Furthermore,

$$\mathbf{a}_2 = \langle \mathbf{q}_1, \mathbf{a}_2 \rangle \mathbf{q}_1 + \mathbf{w} = \langle \mathbf{q}_1, \mathbf{a}_2 \rangle \mathbf{q}_1 + \|\mathbf{w}\| \mathbf{q}_2,$$

and so we obtain (2) by letting $R_{1,2} = \langle \mathbf{q}_1, \mathbf{a}_2 \rangle$ and $R_{2,2} = \|\mathbf{w}\|$.

The general k-th step is similar. For $k = 1, \ldots, m$,

1. let
$$\mathbf{w} = \mathbf{a}_k - \sum_{i=1}^{k-1} \langle \mathbf{q}_i, \mathbf{a}_k \rangle \mathbf{q}_i.$$

2. let
$$\begin{aligned} \mathbf{q}_k &= \mathbf{w}/\|\mathbf{w}\|, \\ R_{i,k} &= \langle \mathbf{q}_i, \mathbf{a}_k \rangle, \qquad i = 1, \ldots, k-1, \\ R_{k,k} &= \|\mathbf{w}\|. \end{aligned}$$

We have now constructed the first m columns of Q and the first m rows of R, and thus we have the intermediate factorization

$$\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_m \end{bmatrix} \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,m} \\ 0 & R_{2,2} & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & R_{m,m} \end{bmatrix}.$$

To complete the factorization, we choose, independently of A, any unit vectors $\mathbf{q}_{m+1}, \ldots, \mathbf{q}_n$ that, together with $\mathbf{q}_1, \ldots, \mathbf{q}_m$ form an orthonormal set $\mathbf{q}_1, \ldots, \mathbf{q}_n$. This can be done using Lemma 2.2 of the book. We make the remaining $n - m$ rows of R zero.

2 Example with $n = m = 3$

We find the QR factorization of

$$A = \begin{bmatrix} 2 & 1 & -3 \\ 0 & 0 & -1 \\ 0 & 1 & 4 \end{bmatrix}.$$

Step 1. Let $\mathbf{w} = \mathbf{a}_1$. Then

$$\mathbf{q}_1 = \frac{\mathbf{w}}{\|\mathbf{w}\|} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

and $R_{1,1} = \|\mathbf{w}\| = 2$.

Step 2. Let

$$\mathbf{w} = \mathbf{a}_2 - \langle \mathbf{q}_1, \mathbf{a}_2 \rangle \mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Then $\mathbf{q}_2 = \mathbf{w}$ and

$$\begin{aligned} R_{1,2} &= \langle \mathbf{q}_1, \mathbf{a}_2 \rangle = 1, \\ R_{2,2} &= \|\mathbf{w}\| = 1. \end{aligned}$$

Step 3. Let

$$\mathbf{w} = \mathbf{a}_3 - \langle \mathbf{q}_1, \mathbf{a}_3 \rangle \mathbf{q}_1 - \langle \mathbf{q}_2, \mathbf{a}_3 \rangle \mathbf{q}_2 = \begin{bmatrix} -3 \\ -1 \\ 4 \end{bmatrix} - (-3) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - 4 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}.$$

Then $\mathbf{q}_3 = \mathbf{w}$ and

$$\begin{aligned} R_{1,3} &= \langle \mathbf{q}_1, \mathbf{a}_3 \rangle = -3, \\ R_{2,3} &= \langle \mathbf{q}_2, \mathbf{a}_3 \rangle = 4, \\ R_{3,3} &= \|\mathbf{w}\| = 1. \end{aligned}$$

Thus,

$$\begin{bmatrix} 2 & 1 & -3 \\ 0 & 0 & -1 \\ 0 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & -3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}.$$

3 Example with $n = 3$, $m = 2$

We find the QR factorization of

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

Using steps 1 and 2 of the previous example we obtain the intermediate factorization

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

To complete the QR factorization, we choose any $\mathbf{q}_3$ such that $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ are orthonormal. For example, we can take $\mathbf{q}_3 = [0, 1, 0]^T$. The QR factorization is then

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

4 Linear least squares

Consider the linear system $A\mathbf{x} = \mathbf{b}$ where A is an $n \times m$ matrix and $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$. Suppose that $n > m$. Then the system is *over-determined.* There are too many equations with respect to the number of unknowns. This situation occurs frequently: we might obtain the equations from n observations and we want to build an m-dimensional linear model with m much smaller than n. In statistics this is known as *linear regression.*

The simplest approach is to seek $\mathbf{x} \in \mathbb{R}^m$ that minimizes the Euclidean norm $\|A\mathbf{x} - \mathbf{b}\|$. This is known as the *least-squares problem.*

Theorem 1 *The vector $\mathbf{x} \in \mathbb{R}^m$ minimizes $\|A\mathbf{x} - \mathbf{b}\|$ if and only if it minimizes $\|\Omega A\mathbf{x} - \Omega\mathbf{b}\|$ for any orthogonal matrix Ω.*

Proof. Since Ω is orthogonal, if $\mathbf{v}$ is any vector in $\mathbb{R}^n$,

$$\|\Omega\mathbf{v}\|^2 = \mathbf{v}^T\Omega^T\Omega\mathbf{v} = \mathbf{v}^T\mathbf{v} = \|\mathbf{v}\|^2.$$

Therefore, for any $\mathbf{x} \in \mathbb{R}^m$,

$$\|\Omega A\mathbf{x} - \Omega\mathbf{b}\| = \|A\mathbf{x} - \mathbf{b}\|.$$

□

Suppose now for simplicity that A has (full) rank m, i.e., its m columns are linearly independent. Let us see how we can use a QR factorization of A to solve the least squares problem. We let $\Omega = Q^T$ in the theorem so that

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \|Q^TA\mathbf{x} - Q^T\mathbf{b}\|^2 = \|R\mathbf{x} - Q^T\mathbf{b}\|^2.$$

Next we write R in block form,

$$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

where R_1 is $m \times m$. R_1 is upper triangular and non-singular. We correspondingly partition the vector $Q^T\mathbf{b}$ into two vectors:

$$Q^T\mathbf{b} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix},$$

where $\mathbf{c}_1$ has length m, and $\mathbf{c}_2$ has length $n - m$. Now we find that

$$\|R\mathbf{x} - Q^T\mathbf{c}\|^2 = \left\| \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} \right\|^2 = \|R_1\mathbf{x} - \mathbf{c}_1\|^2 + \|\mathbf{c}_2\|^2.$$

Thus the solution to the least squares problem is the vector $\mathbf{x} \in \mathbb{R}^m$ that solves

$$R_1\mathbf{x} = \mathbf{c}_1,$$

and the minimum is

$$\min_{\mathbf{x}\in\mathbb{R}^m} \|A\mathbf{x} - \mathbf{c}\| = \|\mathbf{c}_2\|.$$

5 Exercises about orthogonal matrices

Exercise 1 *Show that if Q is an $n \times n$ orthogonal matrix and $\mathbf{x}$ is a vector in $\mathbb{R}^n$, then $\|Q\mathbf{x}\| = \|\mathbf{x}\|$, where $\|\cdot\|$ is the Euclidean norm (2-norm) in $\mathbb{R}^n$.*

Exercise 2 *Show that if Q and P are $n \times n$ orthogonal matrices then PQ is also orthogonal.*

Exercise 3 *Show that the matrix*

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is orthogonal for any $\theta \in \mathbb{R}$. What is its determinant? What is the result $Q\mathbf{x}$ of applying Q to $\mathbf{x} = r[\cos\alpha, \sin\alpha]^T$, for $r \geq 0$ and $\alpha \in \mathbb{R}$?

Exercise 4 *What is the inverse of Q in Exercise 3? What is $Q^{-1}\mathbf{x}$ when $\mathbf{x} = r[\cos\alpha, \sin\alpha]^T$?*

Exercise 5 *If*

$$P = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix}$$

and Q is as in Exercise 3, what is PQ?

Exercise 6 *Is the matrix*

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

orthogonal? What is $Q\mathbf{x}$ when $\mathbf{x} = [x_1, x_2]^T$?

Exercise 7 *Suppose $\mathbf{r}_1, \ldots, \mathbf{r}_n$ and $\mathbf{s}_1, \ldots, \mathbf{s}_n$ are two sets of orthonormal vectors in $\mathbb{R}^n$. Show that the matrix*

$$Q = \begin{bmatrix} \langle \mathbf{r}_1, \mathbf{s}_1 \rangle & \langle \mathbf{r}_1, \mathbf{s}_2 \rangle & \cdots & \langle \mathbf{r}_1, \mathbf{s}_n \rangle \\ \langle \mathbf{r}_2, \mathbf{s}_1 \rangle & \langle \mathbf{r}_2, \mathbf{s}_2 \rangle & \cdots & \langle \mathbf{r}_2, \mathbf{s}_n \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{r}_n, \mathbf{s}_1 \rangle & \langle \mathbf{r}_n, \mathbf{s}_2 \rangle & \cdots & \langle \mathbf{r}_n, \mathbf{s}_n \rangle \end{bmatrix}$$

is orthogonal.