Lecture `linalg6`: QR–factorization: Gram-Schmidt process

November 7, 2022

Summary: Supplemental material on QR–factorization via Gram-Schmidt process, whereas Sauer uses Householder reflections and previous notes describe Givens rotations.

References: T.Sauer, *Numerical Analysis*, Section 4.3, pages 218–229.

Background

There are two versions of the Gram-Schmidt process:[1] classical Gram-Schmidt and modified Gram-Schmidt. We describe the classical algorithm, although the modified process is better behaved on a computer with inexact arithmetic! Assume $A = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n) \in \mathbb{R}^{m\times n}$, with $m \geq n$ and $\mathbf{a}_j \in \mathbb{R}^m$ the jth column of A. We assume a "thin" QR-factorization $A = \hat{Q}\hat{R}$ of the form

$$\overbrace{(\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)}^{A} = \overbrace{(\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n)}^{\hat{Q}} \overbrace{\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}}^{\hat{R}}.$$

Here $\hat{R} \in \mathbb{R}^{n\times n}$ is square and upper-triangular, but, unless $m = n$, $\hat{Q} \in \mathbb{R}^{m\times n}$ is not square, and so cannot be orthogonal. Unless $m = n$, $\hat{Q}^T = \hat{Q}^{-1}$ cannot hold. Nonetheless, even if $m > n$ the matrix $\hat{Q}$ has orthonormal columns, and $\hat{Q}^T\hat{Q} = I_{n\times n}$ (but $\hat{Q}\hat{Q}^T \neq$ identity). Comparing columns of the equation above, we see that

$$\mathbf{a}_j = \sum_{i=1}^{j} r_{ij}\mathbf{q}_i \qquad \text{for } i = 1, \ldots, n.$$

That is, written out explicitly

$$\begin{aligned}
\mathbf{a}_1 &= r_{11}\mathbf{q}_1 \\
\mathbf{a}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 \\
\mathbf{a}_3 &= r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3 \\
&\vdots \\
\mathbf{a}_n &= r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + r_{3n}\mathbf{q}_3 + \cdots r_{nn}\mathbf{q}_n.
\end{aligned}$$

These expansions allow for recursive computation of the numbers r_{ij} and orthonormal vectors $\mathbf{q}_j$. Indeed, we immediately see that $r_{11} = \|\mathbf{a}_1\|$, since $\|\mathbf{q}_1\| = 1$ by hypothesis. Then $\mathbf{q}_1 = \mathbf{a}_1/r_{11}$. Now

[1] According to Wikipedia, Jørgen Pedersen Gram (27 June 1850 – 29 April 1916) was a Danish actuary and mathematician, and Erhard Schmidt (13 January 1876 – 6 December 1959) was a Baltic German mathematician. Gram's work appeared in 1883.

set $\mathbf{v}_2 = \mathbf{a}_2 = r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2$. Since we already know $\mathbf{q}_1$ and since $\mathbf{q}_1^T\mathbf{q}_2 = 0$ by hypothesis, we compute $r_{12} = \mathbf{q}_1^T\mathbf{v}_2$. Therefore, we can compute $\mathbf{v}_2 - r_{12}\mathbf{q}_1 = r_{22}\mathbf{q}_2$. Based on this formula we overwrite $\mathbf{v}_2 \leftarrow \mathbf{v}_2 - r_{12}\mathbf{q}_1$, and then define $r_{22} = \|\mathbf{v}_2\|$ and $\mathbf{q}_2 = \mathbf{v}_2/r_{22}$. Now with $\mathbf{q}_1$ and $\mathbf{q}_2$, we may similarly compute r_{13}, r_{23}, r_{33}, and $\mathbf{q}_3$.

The process just described can be carried out to the end provided all encountered $r_{jj} \neq 0$. The conditions $r_{jj} \neq 0$ are guaranteed if the set $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is linearly independent. Indeed, assume linear independence:

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = \mathbf{0} \implies c_1 = c_2 = \cdots = c_n = 0.$$

Substitution with the above expansions yields

$$c_1(r_{11}\mathbf{q}_1) + c_2(r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2) + \cdots + c_n(r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + \cdots r_{nn}\mathbf{q}_n) = \mathbf{0}.$$

Application of $\mathbf{q}_n^T$ to both sides then yields $r_{nn}c_n = 0$. Since we have assumed $c_n = 0$ must hold necessarily, $r_{nn} \neq 0$. Now, if $r_{nn} \neq 0$ and so $c_n = 0$, then the last equation becomes

$$c_1(r_{11}\mathbf{q}_1) + c_2(r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2) + \cdots + c_{n-1}(r_{1,n-1}\mathbf{q}_1 + r_{2,n-1}\mathbf{q}_2 + \cdots r_{n-1,n-1}\mathbf{q}_{n-1}) = \mathbf{0}.$$

We may now similarly conclude $r_{n-1,n-1} \neq 0$. Proceeding inductively, we find $r_{jj} \neq 0$ for all $j = 1, \ldots, n$ if the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent. The algorithm for classical Gram-Schmidt is given below.

Algorithm 1: Classical Gram-Schmidt algorithm

Data: Linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \in \mathbb{R}^m$
Result: Orthonormal columns $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n \in \mathbb{R}^m$ and an upper-triangular matrix $\hat{R} \in \mathbb{R}^{n\times n}$,
such that $A = \hat{Q}\hat{R}$, where $A = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)$ and $\hat{Q} = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n)$.

```
for j = 1 to n do
    v_j = a_j
    for i = 1 to j − 1 do
        Compute r_ij = q_i^T v_j
        Overwrite v_j ← v_j − r_ij q_i
    end
    r_jj = ‖v_j‖
    q_j = v_j / r_jj
end
```

Example

We wish to solve

$$\begin{pmatrix} 2 & 3 \\ -2 & -6 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \\ 6 \end{pmatrix}$$

in the least squares sense. The matrix is $A = (\mathbf{a}_1, \mathbf{a}_2)$, where

$$\mathbf{a}_1 = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}, \qquad \mathbf{a}_2 = \begin{pmatrix} 3 \\ -6 \\ 0 \end{pmatrix}.$$

We will first construct a "thin decomposition"

$$(\mathbf{a}_1, \mathbf{a}_2) = (\mathbf{q}_1, \mathbf{q}_2)\begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{pmatrix},$$

which is equivalent to the equations

$$\mathbf{a}_1 = r_{11}\mathbf{q}_1, \qquad \mathbf{a}_2 = r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2.$$

From the first equation, $r_{11} = \|\mathbf{a}_1\|_2 = 3$, which yields

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{r_{11}} = \begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{pmatrix}.$$

Rewrite the second equation as $r_{22}\mathbf{q}_2 = \mathbf{a}_2 - r_{12}\mathbf{q}_1$. Using $\mathbf{q}_1^T\mathbf{q}_2 = 0$, we get $r_{12} = \mathbf{q}_1^T\mathbf{a}_2 = 2+4+0 = 6$, and

$$r_{22}\mathbf{q}_2 = \mathbf{a}_2 - (\mathbf{q}_1^T\mathbf{a}_2)\mathbf{q}_1 = \begin{pmatrix} 3 \\ -6 \\ 0 \end{pmatrix} - 6\begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ -2 \end{pmatrix}, \qquad r_{22} = \left\| \begin{pmatrix} -1 \\ -2 \\ -2 \end{pmatrix} \right\|_2 = 3.$$

Finally,

$$\mathbf{q}_2 = \frac{1}{r_{22}}\begin{pmatrix} -1 \\ -2 \\ -2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} \\ -\frac{2}{3} \\ -\frac{2}{3} \end{pmatrix},$$

and our "thin decomposition" is then

$$\begin{pmatrix} 2 & 3 \\ -2 & -6 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{2}{3} & -\frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{pmatrix}\begin{pmatrix} 3 & 6 \\ 0 & 3 \end{pmatrix}.$$

To get a "thick" QR decomposition, where Q is orthogonal, we construct a third unit vector $\mathbf{q}_3$ which is orthogonal to $\mathbf{q}_1$ and $\mathbf{q}_2$. The cross product affords one construction,

$$\mathbf{q}_3 = \mathbf{q}_1 \times \mathbf{q}_2 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \end{vmatrix} = \frac{2}{3}\mathbf{i} + \frac{1}{3}\mathbf{j} - \frac{2}{3}\mathbf{k} = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} 2 & 3 \\ -2 & -6 \\ 1 & 0 \end{pmatrix} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3]\begin{pmatrix} r_{11} & r_{12} \\ 0 & r_{22} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \end{pmatrix}\begin{pmatrix} 3 & 6 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}.$$

Using MATLAB, we find the following decomposition.

```
>> format compact
>> A = [2 3; -2 -6; 1 0];
>> [Q R] = qr(A);
>> rats(A)
ans =
          2           3
         -2          -6
          1           0
>> rats(Q)
ans =
       -2/3        -1/3        -2/3
        2/3        -2/3        -1/3
       -1/3        -2/3         2/3
>> rats(R)
ans =
         -3          -6
          0           3
          0           0
```

This differs slightly from our hand-constructed decomposition, but both are valid QR decompositions. Indeed, since $[\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3]$ is an orthogonal matrix, so is $[-\mathbf{q}_1, \mathbf{q}_2, -\mathbf{q}_3]$.

Now we rewrite the above least squares problem as $QR\mathbf{x} = \mathbf{b}$,

$$\begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \end{pmatrix} \begin{pmatrix} 3 & 6 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \\ 6 \end{pmatrix},$$

or, since the first matrix Q is orthogonal,

$$\begin{pmatrix} 3 & 6 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{pmatrix} \begin{pmatrix} 3 \\ -3 \\ 6 \end{pmatrix} = \begin{pmatrix} 6 \\ -3 \\ -3 \end{pmatrix},$$

that is as $R\mathbf{x} = Q^T\mathbf{b} = \mathbf{w}$. The least squares solution is

$$x_1^{LS} = 4, \qquad x_2^{LS} = -1,$$

and in this case the least square error is

$$\|A\mathbf{x}^{LS} - \mathbf{b}\| = \|R\mathbf{x}^{LS} - \mathbf{w}\| = |w_3| = 3.$$