Gram-Schmidt Process and the QR-decomposition

by Marek Rychlik

March 6, 2009

Gram-Schmidt Process

Let v_i, $i=1,2,...,k$ be a sequence of vectors in $\mathbb{R}^n$. Let us consider the sequence of vector subspaces spanned by the consecutive vectors:

$$\begin{aligned} V_1 &= \text{span}\{v_1\}, \\ V_2 &= \text{span}\{v_1, v_2\}, \\ &\dots \\ V_k &= \text{span}\{v_1, v_{2,} ..., v_k\}. \end{aligned}$$

For simplicity, we assume that the set $\{v_i\}_{i=1}^k$ is linearly independent, although it is important to consider linearly dependent sets, too. The consequence of independence is that:

$$V_1 \subset V_2 \subset ... \subset V_k$$

is a strictly increasing sequence of vector spaces.

We can find a sequence of mutually orthogonal vectors q_j, $j = 1, 2, ..., p$ such that:

$$V_i = \text{span}\{q_1, q_2, ..., q_i\}, \ i = 1, 2, ..., k.$$

That is, the sequence $\{v_1, v_2, ..., v_k\}$ can be replaced with a sequence of orthogonal vectors. Sometimes one assumes that the vectors are normalized, which is easy to achieve as an extra step at the end.

The algorithm

Let $\text{proj}_W(v)$ denote the **orthogonal** projection of a vector v on a vector subspace $W \subseteq \mathbb{R}^n$.
We define the vectors q_i by induction as follows:

$$\begin{aligned} q_1 &= v_1, \\ q_2 &= v_2 - \text{proj}_{V_1}(v_2), \\ &\dots \\ q_i &= v_i - \text{proj}_{V_{i-1}}(v_i), \\ &\dots \\ q_k &= v_k - \text{proj}_{V_{k-1}}(v_k). \end{aligned}$$

The projection on a space V_{i-1} spanned by orthogonal vectors $\{q_1, q_2, \dots, q_{i-1}\}$ can be expressed explicitly as follows:

$$\text{proj}_{V_{i-1}}(v_k) = \sum_{j=1}^{i-1} \frac{\langle v_i, q_j \rangle}{\langle q_j, q_j \rangle} q_j.$$

The angle bracket notation $\langle u, v\rangle$ denotes the dot product, although could be replaced by any bilinear form that is non-degenerate. For the purpose of this article, we assume the standard dot product:

$$\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i = u^T v = \begin{pmatrix} u_1 & u_2 & \dots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

An alternative form of the algorithm

The above version of the algorithm does not produce normalized vectors. However, these three formulas do:

$$\begin{aligned} q_1 &= \frac{1}{\|v_1\|} v_1, \\ \tilde{q}_i &= v_i - \sum_{j=1}^{i-1} \langle v_j, q_j\rangle q_j, \\ q_i &= \frac{1}{\|\tilde{q}_i\|} \tilde{q}_i. \end{aligned}$$

An easy analysis shows that the vectors it produces are the normalized versions of the previous algorithm. Since $\langle q_i, q_i \rangle = 1$ at every step, there is no need to divide by $\langle q_i, q_i \rangle$.

Rewriting the result as QR-decomposition

The results of Gram-Schmidt process may be compactly written down as the QR-decomposition. First, we express v_i through q_i:

$$
\begin{aligned}
v_1 &= r_{11} q_1, \\
v_2 &= r_{12} q_1 + r_{22} q_2, \\
&\dots \\
v_i &= r_{1i} q_1 + r_{2i} q_2 + \dots + r_{ii} q_i, \\
&\dots \\
v_k &= r_{1i} q_1 + r_{2k} q_2 + \dots + r_{kk} q_k.
\end{aligned}
$$

By comparing with the expressions for the projections we find that

$$
r_{ji} = \frac{\langle v_i, q_j \rangle}{\langle q_j, q_j \rangle}, \quad 1 \le i \le k, 1 \le j \le i.
$$

We define matrix A by combining vectors v_j as columns of A:

$$A = \left(\; v_1 | v_2 | \ldots | v_k \; \right)$$

Similarly, we combine vectors q_i into a matrix Q:

$$Q = \left(\; q_1 | q_2 | \ldots | q_k \; \right).$$

The matrix R is formed from coefficients r_{ji}:

$$R = \left(\; r_{ij} \; \right).$$

It is easy to check that:

$$A = Q\,R.$$

The fact that the columns of Q are orthogonal may be expressed as:

$$Q^T Q = \mathrm{diag}(\|q_1\|^2, \|q_2\|^2, \ldots, \|q_k\|^2).$$

Moreover, if we use the alternative version of the algorithm that normalizes vectors q_i, we obtain a more easy to remember equation:

$$Q^T Q = I_k$$

where I_k is the $k \times k$ identity matrix.

A statistics example

Example 1. Let us consider the following problem: Given three normal, mean 0, independent, identically distributed random variables (IID's) X_1, X_2 and X_3. Let $Y_1 = 3\bar{X} = X_1 + X_2 + X_3$ be the sample mean. We ask for two complementary normal variables Y_2 and Y_3 which are independent of Y_1 and of each other.

We define consider three vectors $v_1 = (1, 1, 1)$, $v_2 = (0, 1, 0)$ and $v_3 = (0, 0, 1)$. v_1 is chosen so that $Y_1 = v_1^T X$, where $X = (X_1, X_2, X_3)$. Vectors v_2 and v_3 can be chosen more or less at random, but for the purpose of this example we need them to be simple, so we chose two elements of the standard basis of $\mathbb{R}^3$.

We apply the Gram-Shmidt process to our vectors. We obtain:

$$\begin{aligned}
q_1 &= (1, 1, 1) \\
q_2 &= (0, 1, 0) - \frac{\langle v_2, q_1 \rangle}{\langle q_1, q_1 \rangle} q_1 = (0, 1, 0) - \frac{1}{3}(1, 1, 1) \\
&= \left(-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3} \right). \\
q_3 &= (0, 0, 1) - \frac{\langle v_3, q_1 \rangle}{\langle q_{1,1} \rangle} q_1 - \frac{\langle v_3, q_2 \rangle}{\langle q_2, q_2 \rangle} q_2
\end{aligned}$$

$$
\begin{aligned}
&= (0,0,1) - \frac{1}{3}(1,1,1) - \frac{-\frac{1}{3}}{\left(-\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + \left(-\frac{1}{3}\right)^2}\left(-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3}\right) \\
&= (0,0,1) - \frac{1}{3}(1,1,1) + \frac{1}{2}\left(-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3}\right) \\
&= (0,0,1) - \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) + \left(-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}\right) = \left(-\frac{1}{2}, 0, \frac{1}{2}\right).
\end{aligned}
$$

We instantiate matrices A, Q and R:

$$
\begin{aligned}
Q &= \begin{pmatrix} 1 & -\frac{1}{3} & -\frac{1}{2} \\ 1 & \frac{2}{3} & 0 \\ 1 & -\frac{1}{3} & \frac{1}{2} \end{pmatrix}, \\
R &= \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}, \\
A &= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.
\end{aligned}
$$

It can be verified directly that $A = QR$ and $Q^T Q = \mathrm{diag}\left(3, \frac{2}{3}, \frac{1}{2}\right)$. If we normalize the columns of Q, we must multiply the rows of R by the corresponding norms of the columns of Q, i.e. factors $\sqrt{3}$, $\sqrt{\frac{2}{3}}$ and $\sqrt{\frac{1}{2}}$:

$$
\begin{aligned}
Q &= \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix} \\
R &= \begin{pmatrix} \sqrt{3} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}
\end{aligned}
$$

It is obvious why we avoid normalization in by-hand calculations.

The answer to our statistics problem is that Y_1, Y_2 and Y_3 are defined using the coefficients of q_1, q_2 and q_3 (the normalized variant):

$$
\begin{aligned}
Y_1 &= q_1^T X = \frac{1}{\sqrt{3}} X_1 + \frac{1}{\sqrt{3}} X_2 + \frac{1}{\sqrt{3}} X_3, \\
Y_2 &= q_2^T X = -\frac{1}{\sqrt{6}} X_1 + \frac{\sqrt{2}}{\sqrt{3}} X_2 - \frac{1}{\sqrt{6}} X_2, \\
Y_3 &= q_3^T X = -\frac{1}{\sqrt{2}} X_1 + \frac{1}{\sqrt{2}} X_3.
\end{aligned}
$$

A direct application is to rewrite the Studentized mean $\bar{X}$ as a ratio with independent numerator end denominator:

$$
t = \frac{\bar{X}}{\sqrt{3}\sqrt{s_{\bar{X}}^2}} = \frac{Y_1}{\sqrt{(Y_2^2 + Y_3^2)/2}}.
$$

This is the classical representation of the Student t-statistic as ratio of a normally distributed random variable by the square root of a χ^2-distributed variable with 2 degrees of freedom, divided degrees of freedom. To show this, we need the equation:

$$
\operatorname{Cov}\left(\sum_{i=1}^n c_i X_i, \sum_{i=1}^n d_i X_i \right) = \sum_{i=1}^n c_i d_i = \langle c, d \rangle.
$$

Using this identity we verify that:

$$\mathrm{Cov}(Y_i, Y_j) = \delta_{ij} \quad \text{(Kronecker delta)}.$$

We leave the remaining easy calculations to the reader.
We need to express $s_{\bar{X}}^2$ using Y_2 and Y_3. We have:

$$s_{\bar{X}}^2 = \frac{Z_1^2 + Z_2^2 + Z_3^2}{2}$$

where:

$$\begin{aligned} Z_1 &= X_1 - \frac{1}{3}(X_1 + X_2 + X_3), \\ Z_2 &= X_2 - \frac{1}{3}(X_1 + X_2 + X_3), \\ Z_3 &= X_3 - \frac{1}{3}(X_1 + X_2 + X_3). \end{aligned}$$

We rewrite this as a single matrix equation $Z = BX$ where

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

It is good to understand B on a more conceptual level. We first note that

$$Z_i = X_i - \frac{1}{\sqrt{3}} q_1^T X, \quad i = 1, 2, 3.$$

These equations can be rewritten in matrix form:

$$Z = X - \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} q_1^T X = X - q_1 q_1^T X = (I - q_1 q_1^T) X.$$

For any normalized vector q_1, the matrix $P = q_1 q_1^T$ is the matrix of the linear transformation: the orthogonal projection onto the 1-dimensional space $V_1 = \mathrm{span}\{v_1\}$. (We denoted this transformation proj_{V_1} before.) Indeed, for any vector $v \in \mathbb{R}^3$ we have:

$$Pv = q_1 q_1^T v = q_1 (q_1^T v) = (q_1^T v) q_1 = \langle q_1, v \rangle q_1.$$

Thus $Pv \in V_1$. Also, $v - Pv \perp q_1$ because of the following calculation:

$$\langle v - Pv, q_1 \rangle = \langle v, q_1 \rangle - \langle \langle q_1, v \rangle q_1, q_1 \rangle = \langle v, q_1 \rangle - \langle q_1, v \rangle \langle q_1, q_1 \rangle = 0.$$

In the above calculation, we used the fact that $\langle \cdot , \cdot \rangle$ is a **bilinear form**, i.e. it is linear in each argument. By definition, the bilinearity means that for all scalars $\alpha, \beta \in \mathbb{R}$ and all vectors $v, w, z \in \mathbb{R}^3$:

1. $\langle \alpha v + \beta w, z \rangle = \alpha \langle v, z \rangle + \beta \langle w, z \rangle$.

2. $\langle v, \alpha w + \beta z \rangle = \alpha \langle v, w \rangle + \beta \langle v, z \rangle$.

We also used $\langle q_1, q_1 \rangle = 1$.

The next observation is that if we have an orthogonal basis $\{q_1, \ q_2, \ q_3\}$ of $\mathbb{R}^3$ then the following **partition of unity** holds:

$$I = \sum_{i=1}^{3} q_i q_i^T.$$

(Applying both sides to an arbitrary vector $v \in \mathbb{R}^3$ we obtain a longer version of the above:

$$v = \sum_{i=1}^{3} \langle q_i, v \rangle q_i$$

This expresses the fact that every vector is the sum of its projections onto vectors q_i, $i=1,2,3$. Last two equations are, of course equivalent.)
Thus,

$$B = I - q_1 q_1^T = q_2 q_2^T + q_3 q_3^T$$

is the projection onto span$\{q_2,\ q_3\}$, the orthogonal complement of q_1 (denoted span$\{q_1\}^\perp$).

In order to rewrite $s^2_{\bar{X}}$ in terms of Y_2 and Y_3, we neeed to express $\|Z\|^2$ in terms of Y_j. We observe that $Y = Q^T X$ by definition (equivalent to $Y_j = q_j^T X$). Thus, $X = QY$. Hence $Z = BQY$. We find the product BQ:

$$BQ = (Bq_1|Bq_2|Bq_3) = \begin{pmatrix} 0|q_2|q_3 \end{pmatrix}.$$

(Projection onto span$\{q_2,\ q_3\}$ maps q_1 to the zero vector, and it maps q_2 and q_3 to themselves. Also, a matrix product like BQ can be computed by applying B to columns of Q separately.)

Let $\tilde{Q}$ denote the matrix Q with the first column dropped:

$$Q = \begin{pmatrix} q_2 & q_3 \end{pmatrix}.$$

Thus, $Q = \left(0 | \tilde{Q} \right)$. Let $\tilde{Y} = (Y_2, Y_3)$ denote Y with Y_1 dropped. We have:

$$Z = BQY = \left(0 | \tilde{Q} \right) \begin{pmatrix} Y_1 \\ \tilde{Y} \end{pmatrix} = \tilde{Q}\tilde{Y}$$

Clearly:

$$\|Z\|^2 = \|\tilde{Q}\tilde{Y}\|^2 = (\tilde{Q}\tilde{Y})^T(\tilde{Q}\tilde{Y}) = \tilde{Y}^T(\tilde{Q}^T\tilde{Q})\tilde{Y} = \tilde{Y}^T I_2 \tilde{Y} = \tilde{Y}^T\tilde{Y} = \|\tilde{Y}\|^2.$$

In other words:

$$Z_1^2 + Z_2^2 + Z_3^2 = Y_2^2 + Y_3^2.$$

Finally,

$$s_{\bar{X}}^2 = \frac{Z_1^2 + Z_2^2 + Z_3^2}{2} = \frac{Y_2^2 + Y_3^2}{2}.$$

and

$$t = \frac{Y_1}{\sqrt{(Y_2^2 + Y_3^2)/2}}.$$

Computing QR-decompositions with R

The function `qr` computes the QR-decomposition.

```
> A <- matrix(c(1,1,1,0,1,0,0,0,1),nrow=3,ncol=3)
> A
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    1    1    0
[3,]    1    0    1
> qr <- qr(A)
> qr
$qr
           [,1]       [,2]       [,3]
[1,] -1.7320508 -0.5773503 -0.5773503
[2,]  0.5773503 -0.8164966  0.4082483
[3,]  0.5773503 -0.2588190  0.7071068

$rank
[1] 3

$qraux
[1] 1.5773503 1.9659258 0.7071068
```

```
$pivot
[1] 1 2 3

attr(,"class")
[1] "qr"
```

```
> Q <- qr.Q(qr)
> R <- qr.R(qr)
> Q
```

```
           [,1]       [,2]       [,3]
[1,] -0.5773503  0.4082483 -0.7071068
[2,] -0.5773503 -0.8164966  0.0000000
[3,] -0.5773503  0.4082483  0.7071068
```

```
> R
```

```
          [,1]       [,2]       [,3]
[1,] -1.732051 -0.5773503 -0.5773503
[2,]  0.000000 -0.8164966  0.4082483
[3,]  0.000000  0.0000000  0.7071068
```

```
> A - Q %*% R
              [,1]          [,2]          [,3]
[1,] 1.110223e-16 -1.110223e-16  1.110223e-16
[2,] 0.000000e+00  0.000000e+00 -1.110223e-16
[3,] 0.000000e+00 -5.551115e-17  0.000000e+00
>
```

We notice that there is at least sign difference between our answer calculated by hand and the answer given by the computer. This is an inherent ambiguity of the QR-decomposition, that Q is determined up to flipping the sign of the columns (if we assume $Q^TQ=I$). Let us check our answer further:

```
> Q.byhand = matrix(c(1/sqrt(3),-1/sqrt(6),-
  1/sqrt(2),1/sqrt(3),sqrt(2)/sqrt(3),0,1/sqrt(3),-
  1/sqrt(6),1/sqrt(2)),nrow=3,ncol=3,byrow=T)
> Q.byhand
          [,1]       [,2]       [,3]
[1,] 0.5773503 -0.4082483 -0.7071068
[2,] 0.5773503  0.8164966  0.0000000
```

```
 [3,] 0.5773503 -0.4082483  0.7071068
> R.byhand =
  matrix(c(sqrt(3),1/sqrt(3),1/sqrt(3),0,sqrt(2)/sqrt(3),-
  1/sqrt(6),0,0,1/sqrt(2)),nrow=3,ncol=3,byrow=T)
> R.byhand
         [,1]      [,2]       [,3]
 [1,] 1.732051 0.5773503  0.5773503
 [2,] 0.000000 0.8164966 -0.4082483
 [3,] 0.000000 0.0000000  0.7071068
> crossprod(Q.byhand,Q.byhand)
      [,1] [,2] [,3]
 [1,]    1    0    0
 [2,]    0    1    0
 [3,]    0    0    1
> A-Q.byhand %*% R.byhand
      [,1]          [,2]          [,3]
 [1,]    0  0.000000e+00 -2.220446e-16
 [2,]    0 -2.220446e-16  0.000000e+00
```

```
     [3,]    0  0.000000e+00  0.000000e+00
>
```

So, indeed, the only difference is that of the sign of the first column of Q, and the resulting differences in R. We see by both calculations lead to equation $A = QR$ within the machine round-off error.