A Vector Space Justification of Householder Orthogonalization

Ronald Christensen Professor of Statistics Department of Mathematics and Statistics University of New Mexico

August 28, 2015

Abstract

We demonstrate explicitly that the Householder transformation method of performing a QR decomposition is equivalent to performing the Gram-Schmidt orthonomalization.

KEY WORDS: Gram-Schmidt, Householder transformations, QR decomposition.

1. Introduction

For a standard statistical linear model

$$Y = X\beta + e,$$
 $E(e) = 0,$ $Cov(e) = \sigma^2 I,$

the fundamental statistic is the least squares vector of predicted values, $\hat{Y} \equiv X\hat{\beta}$. This vector is the perpendicular projection of Y into the vector space spanned by the columns of X, often called the column space of X and written C(X). Finding the perpendicular projection essentially requires finding an orthonormal basis for C(X). The most common way of doing that seems to be finding a QR decomposition of X. The purpose of this work is to give a vector (Hilbert) space justification for one of the most common, mathematically stable methods for finding a QR decomposition: the method based on Householder matrices.

Consider an $n \times p$ matrix X with r(X) = r. A QR decomposition of X is a characterization

$$X = QR$$

where $Q_{n \times r}$ has orthonormal columns and $R_{r \times p}$ is upper triangular. The columns of X and Q are the vectors x_j and o_j , respectively, so $\{o_1, \ldots, o_r\}$ in an orthornormal basis for C(X).

Numerically stable approaches to finding a QR decomposition typically find a sequence of orthogonal matrices, Q_1, \ldots, Q_t that upper triangularize X, i.e.,

$$Q_t \cdots Q_1 X = \tilde{R}$$

where \tilde{R} is $n \times p$ with

$$\tilde{R} = \begin{bmatrix} R_{r \times p} \\ 0_{n-r \times p} \end{bmatrix}$$

and R upper triangular. Note that

$$\tilde{Q} \equiv [Q_t \cdots Q_1]'.$$

is an orthogonal matrix because it is the product of orthogonal matrices. The QR decomposition then involves the aforementioned R with Q consisting of the first r columns of \tilde{Q} . The columns of Q are orthonormal and provide a spanning set for C(X), hence give an orthonormal basis. Although this is a nicely terse argument, it is an extremely indirect approach to finding an orthogonal basis. Typically this approach to QR is performed using Householder or Givens transformations.

As discussed in the next section, another method for producing the QR decomposition, and a far more transparent method for producing an orthonormal basis, uses the Gram-Schmidt (G-S) algorithm. Harville (1997) points out that the QR decomposition is unique (for full rank X), from which it follows that any alternative forms of finding QR must ultimately reduce to the G-S algorithm. We demonstrate this equivalence with G-S for the Householder method. (Givens seems less amenable to this purpose.) In fact, our focus is not on finding the QR decomposition but on finding the matrix Q.

2. Gram-Schmidt

The standard theoretical method for orthonormalizing vectors is to use the Gram-Schmidt (G-S) algorithm.

The Gram-Schmidt Theorem.

There exists an orthonormal basis for C(X), say $\{o_1, \ldots, o_r\}$, in which, for $s = 1, \ldots, r$, $\{o_1, \ldots, o_k\}$ is an orthonormal basis for $C(x_1, \ldots, x_s)$ when the space has rank k.

PROOF. Define the o_i s using the G-S Algorithm. Let $||v|| \equiv \sqrt{v'v}$ and let

$$o_{1} = x_{1} / ||x_{1}||,$$

$$z_{s} = x_{s} - \sum_{k=1}^{s-1} o_{k}(o'_{k}x_{s}),$$

$$o_{s} = z_{s} / ||z_{s}||.$$

The proof by induction is standard (when r(X) = p) and omitted. However, if $\{x_1, \ldots, x_p\}$ is not a basis, $z_s = 0$ implies that x_s is a linear function of $\{x_1, \ldots, x_{s-1}\}$. We exclude this z_s from the orthonormal basis being constructed and adjust all subscripts accordingly. Clearly, when constructing a QR decomposition (as opposed to an orthonormal basis), we will be able to write the corresponding x_s as a linear combination of o_1, \ldots, o_k .

The sweep operators used by some statistical regression programs are fairly straight forward applications of this G-S algorithm, also see LaMotte (2014).

Akin to Christensen (2011, Appendix B), G-S essentially defines

$$z_{s+1} = (I - M_k)x_{s+1}$$

where

$$M_k = \sum_{j=1}^k o_j o'_j$$

is the perpendicular projection operator onto $C(x_1, \ldots, x_s)$ which has rank k, i.e., z_{s+1} is the perpendicular projection of x_{s+1} onto the orthogonal complement of $C(x_1, \ldots, x_s)$.

A numerically more stable form of G-S involves taking the product of individual projections operators,

$$z_{s+1} = \left[\prod_{j=1}^{k} (I - o_j o'_j)\right] x_{s+1} = \left[(I - o_k o'_k) \cdots \left[(I - o_2 o'_2)\left[(I - o_1 o'_1) x_{s+1}\right]\right] \cdots\right]$$

Numerical stability is all about the order in which you do things. It is not hard to show that $\prod_{j=1}^{k} (I - o_j o'_j) = (I - M_k)$. This method is not thought to be as stable as triangulation through orthogonal matrices.

3. Householder Transformations

Householder transformations reflect a vector in a (hyper)plane. For a unit vector y with ||y|| = 1 define the Householder transformation as

$$H = I - 2yy'.$$

The reflection in a plane is relative to the plane that is the orthogonal complement of C(y), written $C(y)^{\perp}$. In particular, write any vector x as $x = x_0 + x_1$ with $x_0 \in C(y)$ and $x_1 \perp C(y)$. Then $Hx = -x_0 + x_1$, which is the reflection in the plane.

It is easily seen that each Householder matrix is symmetric and is its own inverse, i.e.,

$$H = H'$$
 and $HH = I$. (1)

In particular, H is an orthogonal matrix.

In the Householder QR decomposition, \tilde{Q} is the product of r specific Householder transformations H_j , j = 1, ..., r so that

$$\tilde{Q} = H_1 \cdots H_r.$$

We now define the matrices H_j . The key idea is to incorporate a previous orthonormal basis, e.g., the columns of the identity matrix $I = [e_1, \dots, e_n]$. We will also use the matrix

$$E_j \equiv \sum_{k=j}^n e_k e'_k$$

which, as j increases, orthogonally projects vectors into smaller and smaller subspaces. (In triangulation discussions of the Householder method, the use of E_j corresponds to looking at smaller and smaller submatrices.) For a collection of nonzero vectors $\{w_1, \ldots, w_r\}$ that we will specify later, define the Householder transformations

$$H_j \equiv I - 2\frac{v_j v'_j}{\|v_j\|^2}, \qquad v_j = w_j - \|w_j\|e_j$$

These particular Householder transforms have a couple of peculiar properties:

$$H_j w_j = ||w_j||e_j, \qquad H_j e_j = (1/||w_j||)w_j.$$
 (2)

These results are not difficult and follow from basic algebraic manipulations. The motivating idea is that H_j rotates the vector w_j into the direction of the orthonormal basis element e_j , an idea that can be used to triangularize a matrix.

Specifically, we define $w_1 \equiv (1/||x_1||)x_1$ and for $j = 1, \ldots, r-1$,

$$w_{j+1} = E_{j+1}H_j \cdots H_1 x_{j+1},$$

with the understanding that, since X has rank r, we will see p-r vectors with $w_{j+1} = 0$ that we exclude from the list. As indicated from our later proof of equivalence with G-S, $w_{j+1} = 0$ indicates $x_{j+1} \in C(x_1, \ldots, x_j)$. Moreover, the very definition of v_{j+1} requires $w_{j+1} \neq 0$. Note that the matrices H_{j+1} are defined recursively.

It is important to show that for k < j,

$$H_j e_k = e_k. \tag{3}$$

This follows directly from orthonormality of the e_i s, the fact that

$$e'_{k}v_{j} = e'_{k}[w_{j} - ||w_{j}||e_{j}] = e'_{k}E_{j}H_{j-1}\cdots H_{1}x_{j} - e'_{k}e_{j}(||w_{j}||) = 0 - 0$$

and that $H_j e_k = I e_k - v_j v'_j e_k (2/||v_j||^2)$. Moreover, it follows immediately from (3) that

$$H_1 \cdots H_j e_k = H_1 \cdots H_k e_k. \tag{4}$$

We can now define our orthonormal basis as

$$o_j \equiv H_1 \cdots H_j e_j = H_1 \cdots H_r e_j,$$

for j = 1, ..., r. It is easy to see that these vectors are orthornormal. Using equations (1)

$$o'_i o_j = [H_1 \cdots H_r e_i]' [H_1 \cdots H_r e_j] = e_i H_r \cdots H_1 H_1 \cdots H_r e_j = e'_i e_j = \begin{cases} 1 & i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

The problem is showing that, in the full rank case say, o_s is in the space spanned by x_1, \ldots, x_s , $s = 1, \ldots, p$. By definition and equations (2), o_1 is in the space spanned by x_1 . The key fact in an induction proof is that $H_1 \cdots H_j e_j$ is equivalent to performing the G-S algorithm on x_j .

We demonstrate for j = 1, ..., r - 1 that $o_{j+1} \equiv H_1 \cdots H_{j+1} e_{j+1}$ is the same as the inductive step of G-S. By (2),

$$o_{j+1} \equiv H_1 \cdots H_{j+1} e_{j+1} = H_1 \cdots H_j w_{j+1} (1/||w_{j+1}||).$$

We will show that

$$H_1 \cdots H_j w_{j+1} = z_{j+1}$$

where z_{j+1} is defined in the G-S algorithm. In G-S, this vector would be normalized by dividing it by $||H_1 \cdots H_j w_{j+1}||$ but in the Householder method the normalization occurred through dividing w_{j+1} by $||w_{j+1}||$. These normalizations agree because $||w_{j+1}|| = ||H_1 \cdots H_j w_{j+1}||$, since Householder transformations are rotations that do not change a vector's length.

Starting with the definition of w_{j+1} and assuming that o_1, \ldots, o_j are appropriate orthonormal vectors

$$H_1 \cdots H_j w_{j+1} = H_1 \cdots H_j E_{j+1} H_j \cdots H_1 x_{j+1}$$

$$= H_1 \cdots H_j \left(I - \sum_{k=1}^j e_k e'_k \right) H_j \cdots H_1 x_{j+1}$$

= $H_1 \cdots H_j H_j \cdots H_1 x_{j+1} - \sum_{k=1}^j H_1 \cdots H_j e_k e'_k H_j \cdots H_1 x_{j+1}$
= $x_{j+1} - \sum_{k=1}^j o_k o'_k x_{j+1}.$

This is exactly the vector z_{j+1} from the G-S algorithm. Normalizing it gives o_{j+1} .

References

- Christensen, Ronald (2011). Plane Answers to Complex Questions: The Theory of Linear Models, 4th Edition. Springer-Verlag, New York.
- Harville, David A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.
- LaMotte, Lynn Roy (2014). The Gram-Schmidt Construction as a Basis for Linear Models, The American Statistician, 68, 52-55.