ORTHOGONALIZATION-TRIANGULARIZATION METHODS IN STATISTICAL COMPUTATIONS

Del T. Scott, Brigham Young University

G. Rex Bryce, Brigham Young University

David M. Allen, University of Kentucky

ABSTRACT

Procedures for reducing a data matrix to triangular form using orthogonal transformations are presented, e.g., Householder, Givens, and Modified Gram-Schmidt. Using small numerical examples these procedures are compared to procedures operating on normal equations. We show how an analysis of variance can be constructed from the triangular reduction of the data matrix. Procedures for calculating sums of squares, degrees of freedom, and expected mean squares are presented. These procedures apply even with mixed models and missing data. It is demonstrated that all statistics needed for inference on linear combinations of parameters of a linear model may be calculated from the triangular reduction of the data matrix. Also included is a test for estimability. We also demonstrate that if the computations are done properly some inference is warranted even when the X matrix is ill-conditioned.

Keywords: Orthogonalization, triangular reduction, Householder, Givens, Gram-Schmidt, Cholesky, Analysis of Variance, Expected Mean Squares, Linear Combinations, triangular reduction, estimability.

1.0 INTRODUCTION

It has long been recognized that the triangularization of the matrix of coefficients of a set of simultaneous equations is a simple yet efficient means of obtaining a solution. Indeed, the author of the Chui-chang suan-shu or Nine Chapters on the Mathematical Art written in 250 BC in China, used the technique although a general systematic method of approach is not given (Boyer, 1968). The first systemization of the technique is apparently due to Gauss [1823].

Further systemization came through those who were using the technique in the practical setting of Geodesy. M. H. Doolittle [1878] published in the U.S. Coast and Geodetic report a method which after some slight procedural changes (Dwyer, 1941) became the abbreviated Doolittle method. Benoit [1924] attributes another method to a Commander Cholesky, a commander of artillery in the Geographic Service of the French army who was killed early in the first world war. Variants of this method which has come to be known as the Cholesky factorization or the square root method have been independently developed by several others (e.g., Schur 1917, Banachiewicz 1937, Dwyer 1941). Dwyer [1941] seems to be the first to have supplied a formal proof of the validity of the techniques and indicated their usefulness beyond the solution of normal equations.

The next major developments seem to have been by numerical analysts. Recognizing the numerical inaccuracies inherent in sums of squares and crossproducts they proposed methods of obtaining solutions directly from the original observational equations rather than the more compact normal equations. This required that an orthogonal basis for the matrix of coefficients be found (perhaps only implicitly), along with an upper triangular matrix. Three basic ways of computing this orthogonal-triangular decomposition resulted: elimination (the Gram-Schmidt method), reflection (the Householder transformation; Householder, 1953) and rotation (the Givens transformation; Givens, 1954; Gentleman, 1972). Stewart [1973] gives a complete discussion of these methods.

However, it was up to the statistician to really plumb the depths of the triangulation and discover its many and varied uses. Dwyer [1941,
1944, 1945, 1951] discusses its use in correlation and regression analysis, Lucas [1950] and Gaylord, Lucas and Anderson [1970] discuss applications to expected mean squares and Rohde and Harvey [1965] present methods for estimating linear combinations of parameters along with their variances. In addition Graybill [1969, 1976] and Seber [1977] have made use of these techniques in texts dealing with the theory of linear models. The purpose of this paper is to present some of the statistical uses to which the triangular decomposition can be put. The uses discussed do not exhaust the possibilities but represent those the authors have found most beneficial in their own work.

1.1 Preliminaries and Notation

Define the orthogonal-triangular decomposition of an $n \times p$ matrix $X$ with $\text{rank}(X) = r \leq p$ as

$$X = QT$$  \hspace{1cm} (1.1)

where $Q$ is an orthonormal basis for $X$, i.e., $Q'Q = I$ an identity matrix with $p - r$ zero's on the diagonal and $T$ is a $p \times p$ upper triangular matrix. We shall consider the matrix $Q$ as a triangularization operator since

$$Q'X = Q'T = T.$$  \hspace{1cm} (1.2)

We shall assume that the $Q$-operator as typified in (1.2) is any one of the class of algorithms which operate directly on the columns of $X$, i.e., Gram-Schmidt, Householder or Givens.

The definition of the $Q$-operator can be broadened to include procedures which yield an upper triangular matrix from the normal equations, e.g., Choleski or square root method, the Doolittle method, etc. If we define

$$Q = XR'$$  \hspace{1cm} (1.3)

then

$$Q'X = R \begin{bmatrix} X'X \end{bmatrix} = T$$  \hspace{1cm} (1.4)

where $R$ is defined as an "inverse" of $T$. Thus in this paper references to a $Q$-operator will refer to any of the algorithms which yield an upper triangular matrix whether from a rectangular system (1.2) or from a square symmetric system (1.4).

Since the matrix $T$ will in general be less than full rank, it will be useful to define a generalized inverse of $T$. Without loss of generality, assume that the null rows of $T$ are the last $p-r$ rows and write

$$T = \begin{bmatrix} T_1 & T_{12} \\ 0 & 0 \end{bmatrix}.$$  \hspace{1cm} (1.5)

Following the suggestion of Allen [1974] define

$$T^g = R' = \begin{bmatrix} r_1^{-1} & r_1^{-1}T_{12} \\ 0 & 0 \end{bmatrix}$$  \hspace{1cm} (1.6)

so that $TR' = I$ and

$$R'T = \begin{bmatrix} 1 & T_1^{-1}T_{12} \\ 0 & 0 \end{bmatrix}.$$  \hspace{1cm} (1.7)

It follows from (1.1) that $X'X = T'T$ and thus, that $R'R$ is a generalized inverse of $X'X$ from which it follows that $H = (X'X)^g(X'X) = R'T$. Of course, for $X$ of full rank $R' = T^{-1}$.

We note that $R'$ contains explicit information on the nature of the redundancy in $X$. Suppose $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ where $X_2 = X_1L$ for some matrix $L$. It follows that $Q = \begin{bmatrix} Q_1 & 0 \end{bmatrix}$ and from (1.3) we have $T_1 = Q_1X_1$ and

$$T_{12} = Q_1^TX_2 = Q_1^TX_1L = T_1L.$$  \hspace{1cm} (1.8)

Since $T_1$ is of full rank $L = T_1^{-1}T_{12}$. But $L$ is just a partition of (1.7) which can be computed