More Accurate Linear Least Squares

RICHARD E. VON HOLDT†

The Linear Least-Squares Problem

Let A be a given matrix of m rows and n columns, $(m \geq n)$, so that $AW=0$ implies $W=0$ (*i.e.*, the column vectors of A constitute a linearly independent set), and let Y be a given m-dimensional vector. We seek an n-dimensional vector $\tilde{X}$, so that $|R(\tilde{X})|^2$ is the minimum value of $|R(X)|^2$ where

$$R(X) = AX - Y. \tag{1}$$

A Geometric Derivation of the Classical Solution to the Linear Least-Squares Problem

Let S be that subspace of m-dimensional Euclidean space which is spanned by the column vectors of A. Then for arbitrary X, AX is a vector in S, and $R(X)$ is a vector with initial point at the terminal point of Y and terminal point in S.

Let $A\tilde{X}$ be the orthogonal projection of Y onto S. Then

$$R(\tilde{X}) = A\tilde{X} - Y, \tag{2}$$

is orthogonal to S, or

$$A^T R(\tilde{X}) = 0. \tag{3}$$

For arbitrary X, we have

$$\begin{aligned} R(X) &= [R(X) - R(\tilde{X})] + R(\tilde{X}) \\ &= A(X - \tilde{X}) + R(\tilde{X}). \end{aligned} \tag{4}$$

From (3) and (4), for arbitrary X, we have

$$|R(X)|^2 = |A(X - \tilde{X})|^2 + |R(\tilde{X})|^2 \geq |R(\tilde{X})|^2.$$

Thus $|R(\tilde{X})|^2$ is the minimum value of $|R(X)|^2$ for arbitrary X. Substituting (2) into (3), $\tilde{X}$ must satisfy the relation:

$$A^T A \tilde{X} = A^T Y. \tag{5}$$

From the hypothesis on A, we have

$$A^T A W = 0 \rightarrow W^T A^T A W \equiv |AW|^2 = 0 \rightarrow W = 0. \tag{6}$$

Hence $A^T A$ is a nonsingular $n \times n$ matrix and (5) has a unique solution.

The Solution of (5) by Diagonal Pivots

Let M_k, $(k=1, 2, \cdots, n)$, be the matrix obtained from A by deleting all but the first k columns of A. Then the columns of M_k form a linearly independent set and by the argument of (6), $M_k^T M_k$ is a nonsingular matrix, $(k=1, 2, \cdots, n)$.

† Lawrence Rad. Lab., University of California, Livermore, Calif.

Let P_1 be the determinant of $M_1^T M_1$ and P_k be the determinant of $M_k^T M_k$ divided by the determinant of $M_{k-1}^T M_{k-1}$, $(k=2, 3, \cdots, n)$. Then

$$P_k \neq 0, \qquad (k = 1, 2, \cdots, n). \tag{7}$$

Using the diagonal elements, in increasing order, as pivots, and combining proper multiples of each row into all following rows to produce zeros below the pivot elements in the column containing the pivots in (5) does not change the value of any of the minors of $A^T A$ formed by deleting all but the first k rows and all but the first k columns of $A^T A$. Thus this process of replacing (5) by an equivalent upper-triangular system of equations yields the successive nonzero pivots:

$$P_1, P_2, \cdots, P_n. \tag{8}$$

The above described process is equivalent to premultiplying both sides of (5) by a lower-triangular matrix L, with unit diagonal elements, yielding

$$LA^T A \tilde{X} = LA^T Y \tag{9}$$

and this upper-triangular system is solved by back substitution.

The Solution of (5) by Orthogonalization

Using the columns of A, in increasing order, as pivot columns, and combining proper multiples of each pivot column into all following columns so that the resulting columns are orthogonal to the pivot column, replaces the columns of A by an orthogonal basis for S. Let the matrix which results be denoted by B. Then

$$B = AU \tag{10}$$

where U is an upper-triangular matrix with unit-diagonal elements, and furthermore

$$B^T B = D, \tag{11}$$

where D is a diagonal matrix.

Premultiplying (5) by U^T and replacing $\tilde{X}$ by $UU^{-1}\tilde{X}$, we have

$$DU^{-1}\tilde{X} = B^T Y \tag{12}$$

and from this,

$$\tilde{X} = UD^{-1}B^T Y. \tag{13}$$

Comparison of Methods

Since both $LA^T A$ and L^T are upper-triangular matrices, their product, $LA^T A L^T$, is also upper triangular, besides being symmetric, and is therefore a diagonal matrix. Thus AL^T is a matrix of mutually orthogonal columns and since L^T is upper-triangular with unit-diagonal elements,

$$L^T = U. \tag{14}$$

Again, since L^T is upper-triangular with unit-diagonal elements, the diagonal elements of LA^TA and $D = U^TA^TAU = LA^TAL^T$ are identical, or

$$|B_k|^2 = P_k, \qquad (k = 1, 2, \cdots, n). \tag{15}$$

Let A_k and B_k be the kth columns of A and B, respectively, and let ϵ_k be a scala, defined by

$$|B_k| = \epsilon_k |A_k|, \qquad (k = 1, 2, \cdots, n). \tag{16}$$

Since B_k is a projection of A_k, we have

$$0 < \epsilon_k \leq 1, \qquad (k = 1, 2, \cdots, n), \tag{17}$$

and ϵ_k is a measure of the figure loss encountered in constructing B_k from A_k by orthogonalization, and there is no further figure loss by cancellation in computing,

$$P_k = |B_k|^2, \qquad (k = 1, 2, \cdots, n). \tag{18}$$

In the method of diagonal pivots, $|B_k|^2$ is formed from $|A_k|^2$ by repeated subtractions, and since

$$|B_k|^2 = \epsilon_k^2 |A_k|^2, \tag{19}$$

our measure of the figure loss in this method is ϵ_k^2. Thus the method of orthogonalization has half the figure loss of the method of diagonal pivots.

$R(\tilde{X})$ as a By-Product of Orthogonalization

Let Z_k, $(k=1, 2, \cdots, n+1)$ be defined by

$$Z_1 = Y \tag{20}$$

and

$$Z_{k+1} = Y - \sum_{i=1}^{k} B_j(B_j^TZ_j)/(B_j^TB_j) \quad (k = 1, 2, \cdots, n). \tag{21}$$

Then

$$B_{k+1}^TZ_{k+1} = B_{k+1}^TY, \quad (k = 1, 2, \cdots, n-1), \tag{22}$$

since the columns of B are mutually orthogonal. Setting $k=n$ in (21) and using (22), we have

$$Z_{n+1} = Y - \sum_{j=1}^{n} B_j(B_j^TY)/(B_j^TB_j). \tag{23}$$

Thus Z_{n+1} is the component of Y orthogonal of S, or

$$Z_{n+1} = -R(\tilde{X}). \tag{24}$$

The Inverse of A^TA After the Application of the Method of Orthogonalization

From (10), (11), and (14), we have

$$LA^TAL^T = D. \tag{25}$$

Since L and L^T are nonsingular,

$$A^TA = L^{-1}D(L^T)^{-1} \tag{26}$$

and

$$(A^TA)^{-1} = L^TD^{-1}L. \tag{27}$$

Since the calculation of D by the method of orthogonalization has half the figure loss of the method of diagonal pivots, the former method using (27) yields a computationally more accurate inverse. The elements of this inverse matrix are useful to statisticians in the Theory of Error Analysis.

Details of the Method of Orthogonalization

The matrix $U=L^T$ need not be formed explicitly except in the evaluation of $(A^TA)^{-1}$ by (27).

Let U_k be the matrix which is the $n\times n$ identity except for the elements in the kth row to the right of the diagonal. These elements are the multiples of the kth column of B which are added to the corresponding following columns to yield new following columns, which are orthogonal to the kth column. Then

$$U = U_1U_2 \cdots U_{n-1}. \tag{28}$$

When: 1) the nonidentity elements of U_k have been formed and used to orthogonalize the following columns and Z_k to B_k; 2) $B_k^TY = B_k^TZ_k$ has been formed; and 3) $|B_k|^2$ has been formed, then B_i is no longer needed and the storage cells used for B_i are now available for storing the nonidentity elements of U_k and the scalar $B_k^TY = (B^TY)_k$. Repeating this process until $k=n$, $\tilde{X}$ is evaluated by

$$\tilde{X} = U_1U_2 \cdots U_{n-1}D^{-1}(B^TZ) \tag{29}$$

where we take advantage of all the known zero elements of the matrices involved in performing the indicated matrix premultiplications.

For calculation of $(A^TA)^{-1}$, we have from (27) and (14)

$$(A^TA)^{-1} = (L_{n-1}L_{n-2} \cdots L_2L_1)^TD^{-1}(L_{n-1}L_{n-2} \cdots L_2L_1). \tag{30}$$

The nonidentity elements of the product

$$L_k(L_{k-1}L_{k-2} \cdots L_1)$$

may be stored in the locations occupied by the nonidentity elements of the two factors, $(k=2, \cdots, n-1)$. Having formed L, the diagonal and subdiagonal elements of $(A^TA)^{-1}$ can be formed and stored in the locations occupied by D and L.

Conclusion

Although the number of operations involved is greater in the method of orthogonalization than in the method of diagonal pivots, the increased accuracy is well worth the time and effort. It is to be noted that the method of orthogonalization for weighted polynomial fitting is equivalent to forming a set of weighted orthogonal polynomials, fitting the data to these polynomials, and reducing the combination of these polynomials to a single polynomial in the manner of Tchebycheff.