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 \( X \), so that \( R(X) \) is the minimum value of \( \| R(X) \| \) where
\[
R(X) = AX - Y.
\]

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 \( AX \) be the orthogonal projection of \( Y \) onto \( S \). Then
\[
R(X) = AX - Y,
\]
is orthogonal to \( S \), or
\[
A^T R(X) = 0.
\]

For arbitrary \( X \), we have
\[
R(X) = [R(X) - R(\hat{X})] + R(\hat{X}) = A(X - \hat{X}) + R(\hat{X}).
\]
From (3) and (4), for arbitrary \( X \), we have
\[
| R(X) |^2 = | A(X - \hat{X}) |^2 + | R(\hat{X}) |^2 \geq | R(\hat{X}) |^2.
\]
Thus \( | R(X) |^2 \) is the minimum value of \( | R(X) |^2 \) for arbitrary \( X \). Substituting (2) into (3), \( \hat{X} \) must satisfy the relation:
\[
A^T AX = A^T Y.
\]
From the hypothesis on \( A \), we have
\[
A^T AW = 0 \Rightarrow W^T A^T AW = | AW |^2 = 0 \Rightarrow W = 0.
\]
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 \( M_k \), \((k = 1, 2, \ldots, n)\), 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, \((k = 1, 2, \ldots, n)\).

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 \( M_{k-1}^T M_{k-1} \), \((k = 2, 3, \ldots, n)\). Then
\[
P_k \neq 0, \quad (k = 1, 2, \ldots, n).
\]

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:
\[
P_1, P_2, \ldots, P_n.
\]

The above described process is equivalent to premultiplying both sides of (5) by a lower-triangular matrix \( L \), with unit diagonal elements, yielding
\[
LA^T AX = LA^T Y
\]
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
\[
B = AU
\]
where \( U \) is an upper-triangular matrix with unit-diagonal elements, and furthermore
\[
B^T B = D,
\]
where \( D \) is a diagonal matrix.

Premultiplying (5) by \( U^T \) and replacing \( \hat{X} \) by \( UU^{-1}\hat{X} \), we have
\[
D U^{-1} \hat{X} = B^T Y
\]
and from this,
\[
\hat{X} = U D^{-1} B^T Y.
\]

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 \( A L^T \) 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 \( L^TA \) and \( D = U^TAU = L^TA^TAL^T \) are identical, or
\[
|B_k|^2 = p_k, \quad (k = 1, 2, \ldots, n).  \tag{15}
\]

Let \( A_k \) and \( B_k \) be the \( k \)-th columns of \( A \) and \( B \), respectively, and let \( \epsilon_k \) be a scalar, defined by
\[
|B_k| = \epsilon_k |A_k|, \quad (k = 1, 2, \ldots, n).  \tag{16}
\]

Since \( B_k \) is a projection of \( A_k \), we have
\[
0 < \epsilon_k \leq 1, \quad (k = 1, 2, \ldots, n),  \tag{17}
\]
and \( \epsilon_k \) 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,
\[
P_k = |B_k|^2, \quad (k = 1, 2, \ldots, n).  \tag{18}
\]

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 \( \epsilon_k^2 \). Thus the method of orthogonalization has half the figure loss of the method of diagonal pivots.

### R(\(X\)) as a By-Product of Orthogonalization

Let \( Z_k, (k = 1, 2, \ldots, n+1) \) be defined by
\[
Z_1 = Y \tag{20}
\]
and
\[
Z_{k+1} = Y - \sum_{i=1}^{k} B_i(B_i^T Z_i)/(B_i^T B_i), \quad (k = 1, 2, \ldots, n). \tag{21}
\]

Then
\[
B_{k+1}^T Z_{k+1} = B_{k+1}^T Y, \quad (k = 1, 2, \ldots, n-1), \tag{22}
\]
since the columns of \( B \) are mutually orthogonal. Setting \( k = n \) in (21) and using (22), we have
\[
Z_{n+1} = Y - \sum_{j=1}^{n} B_j(B_j^T Y)/(B_j^T B_j). \tag{23}
\]

Thus \( Z_{n+1} \) is the component of \( Y \) orthogonal of \( S \), or
\[
Z_{n+1} = -R(\bar{X}). \tag{24}
\]

### The Inverse of \( A^T A \) After the Application of the Method of Orthogonalization

From (10), (11), and (14), we have
\[
L^T A^T A L^T = D. \tag{25}
\]

Since \( L \) and \( L^T \) are nonsingular,
\[
(A^T A)^{-1} = L^T D^{-1} L. \tag{27}
\]

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

Let \( U_k \) be the matrix which is the \( n \times n \) identity except for the elements in the \( k \)-th row to the right of the diagonal. These elements are the multiples of the \( k \)-th column of \( B \) which are added to the corresponding following columns to yield new following columns, which are orthogonal to the \( k \)-th column. Then
\[
U = U_1 U_2 \cdots U_{n-1}. \tag{28}
\]

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) \( B_k^T Y = B_k^T Z_k \) has been formed; and 3) \( |B_k|^2 \) has been formed, then \( Bi \) is no longer needed and the storage cells used for \( B_k \) are now available for storing the nonidentity elements of \( U_k \) and the scalar \( B_k^T Y = (B_k^T Y)_k \). Repeating this process until \( k = n \), \( \bar{X} \) is evaluated by
\[
\bar{X} = U_1 U_2 \cdots U_{n-1} D^{-1}(B_k^T Z_k) \tag{29}
\]
where we take advantage of all the known zero elements of the matrices involved in performing the indicated matrix premultiplications.

For calculation of \((A^T A)^{-1}\), we have from (27) and (14)
\[
(A^T A)^{-1} = (L_{n-1} L_{n-2} \cdots L_2 L_1)^T D^{-1}(L_{n-1} L_{n-2} \cdots L_2 L_1). \tag{30}
\]

The nonidentity elements of the product
\[
L_k(L_{k-1} L_{k-2} \cdots L_i)
\]
may be stored in the locations occupied by the nonidentity elements of the two factors, \( (k = 2, \ldots, n-1) \). Having formed \( L_k \), the diagonal and subdiagonal elements of \((A^T A)^{-1}\) 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.
The CORDIC Computing Technique

JACK VOLDER†

The "COordinate Rotation DIGital Computer" computing technique can be used to solve, in one computing operation and with equal speed, the relationships involved in plane coordinate rotation; conversion from rectangular to polar coordinates; multiplication; division; or the conversion between a binary- and a mixed-radix system.

The CORDIC computer can be described as an entire transfer computer with a special serial arithmetic unit, consisting of 3 shift registers, 3 adder-subtractors, and special interconnections. The arithmetic unit performs a sequence of simultaneous conditional additions or subtractions of shifted numbers to each register. This performance is similar to a division operation in a conventional computer.

Only the trigonometric algorithms used in the CORDIC computing technique will be covered in this paper. These algorithms are suitable only for use with a binary code. This fact possibly accounts for their late appearance as a numerical computing technique. Matrix theory, complex-number theory, or trigonometric identities can be used to prove rigorously these algorithms. However, to help give a more intuitive and pictorial understanding of the basic technique, plane trigonometry and analytical geometry are used in this explanation whenever possible.

First, consider two given coordinate components \( Y_i \) and \( X_i \) in the plane coordinate system shown in Fig. 1.

\[
Y_i = R_i \sin \theta_i \tag{1}
\]

\[
X_i = R_i \cos \theta_i \tag{2}
\]

With a very simple control of an arithmetic unit operating in a binary code, the sign of a number can be changed and/or the number can be divided by a power of two. Thus, if it is assumed that the numerical values of \( Y_i \) and \( X_i \) are available, the numerical values of both coordinates of one of the proportional quadrature vectors, \( R_i' \), can be easily obtained.

\[
Y_i' = 2^{-j}X_i \tag{3}
\]

\[
X_i' = -2^{-j}Y_i \tag{4}
\]

where \( j \) is a positive integer or zero.

The vector addition of \( R_i' \) to \( R_i \) by the algebraic addition of corresponding components, produces the following relationships:

\[
Y_{i+1} = \sqrt{1 + 2^{-2j}} R_i \sin (\theta_i + \tan^{-1} 2^{-j}) = Y_i + 2^{-j}X_i \tag{5}
\]

\[
X_{i+1} = \sqrt{1 + 2^{-2j}} R_i \cos (\theta_i + \tan^{-1} 2^{-j}) = X_i - 2^{-j}Y_i \tag{6}
\]

\[
R_{i+1} = \sqrt{1 + 2^{-2j}} R_i. \tag{7}
\]

Likewise, the addition of the other proportional quadrature vector at \( \theta - 90^\circ \) to the vector \( R_i \) produces the following relationships:

\[
Y_{i+1} = \sqrt{1 + 2^{-2j}} R_i \sin (\theta_i - \tan^{-1} 2^{-j}) = Y_i - 2^{-j}X_i \tag{8}
\]

\[
X_{i+1} = \sqrt{1 + 2^{-2j}} R_i \cos (\theta_i - \tan^{-1} 2^{-j}) = X_i + 2^{-j}Y_i \tag{9}
\]

\[
R_{i+1} = \sqrt{1 + 2^{-2j}} R_i. \tag{10}
\]

If the numerical values of the components \( Y_i \) and \( X_i \) are available, either of the two sets of components \( Y_{i+1} \) and \( X_{i+1} \) may be obtained in one word-addition time with a special arithmetic unit (as shown in Fig. 2) operating serially in a binary code.

This particular operation of simultaneously adding (or subtracting) the shifted \( X \) value to \( Y \) and subtracting (or adding) the shifted \( Y \) value to \( X \) is termed "cross addition."