SAS® Technical Report R-104

A Simplified Algorithm for the W-Transformation in Variance Component Estimation

5983

SAS Institute Inc.

A Simplified Algorithm for the W-Transformation in Variance Component Estimation

by

J. H. Goodnight W. J. Hemmerle

SAS® Technical Report R-104

Abstract

The W transformation is needed at each step in the maximum likelihood or restricted maximum likelihood procedure for estimation of the parameters of the mixed A.O.V. model. This paper develops an efficient algorithm for computing the W transformation needing only about a dozen lines of Fortran or PL/1 code.

SAS Institute Inc.
SAS Circle ☐ Box 8000
Cary, NC 27512-8000

The correct bibliographic citation for this technical report is as follows: SAS Institute Inc., SAS® Technical Report R-104, A Simplified Algorithm for the W-Transformation in Variance Component Estimation, Cary, NC: SAS Institute Inc., 1978.

SAS® Technical Report R-104, A Simplified Algorithm for the W-Transformation in Variance Component Estimation

1. Introduction

The W transformation was suggested by Hemmerle and Hartley [3] for maximum likelihood (ML) estimation of the parameters of the mixed analysis of variance model. This transformation was later applied by Corbeil and Searle [1] to obtain restricted maximum likelihood (REML) estimators and by Liu and Senturia [8] for MINQUE variance component estimators. Hemmerle and Downs [5] consider its use with the mixed model when error variances are unequal. Formulas were developed by Thompson [9] to reduce the computational burden of the transformation by up to a factor of four. Hemmerle and Lorens [4] achieved the same computational savings through an "in-place" algorithm which considered non-negative constraints on variance component estimators as an integral part. Jennrich and Sampson [6], [7] have also developed very effective algorithms for both ML and REML estimation which utilize the W transformation.

For ML or REML estimation, the W transformation is needed at each iterative step where it consumes a substantial amount of the computations for the step. Consequently, it is important that the algorithm for this transformation be as efficient as possible. In this paper, we develop a simplified algorithm for the W transformation which has essentially the same computational efficiency and storage economy as the Hemmerle and Lorenz algorithm with two decided advantages:

a) The simplified algorithm requires only about a dozen lines of FORTRAN or PL/1 code; and

b) No row or column transpositions of the initial W_o matrix are ever needed.

In what follows, we restrict the discussion to ML estimation and consider the model

$$Y = X\alpha + \varepsilon \tag{1}$$

where ε has a multivariate normal distribution with variance-covariance matrix $\sigma^2 H$ where

$$H = I_n + \sum_{i=1}^{c} \gamma_i U_i U_i' , \tag{2}$$

X is a known nxk matrix, U_i are known n x m_i matrices associated with the variance components $\sigma^2\gamma_i$, Y is an nx1 vector of observations, and α is a kx1 vector of fixed but unknown parameters. Letting

$$m = \sum_{i=1}^{c} m_i$$

and $V = [U_1|U_2|\cdots|U_c]$, the W-transformation maps the matrix

$$W_o = \left[\begin{array}{c|c|c} V'V & V'X & V'Y \\ \hline X'V & X'X & X'Y \\ \hline Y'V & Y'X & Y'Y \end{array}\right] \tag{3}$$

into the matrix

$$W = \left[\begin{array}{c|c|c} V'H^{-1}V & V'H^{-1}X & V'H^{-1}Y \\ \hline X'H^{-1}V & X'H^{-1}X & X'H^{-1}Y \\ \hline Y'H^{-1}V & Y'H^{-1}X & Y'H^{-1}Y \end{array}\right] \tag{4}$$

Various functions of the elements of W are then employed in the current ML step to arrive at new parameter estimates.

The algorithm presented in [4] computes W in place, using only the elements of the upper triangle of W_o, using approximately $m(m+k+1)^2/2$ operations (multiplications and divisions); however, there are several separate types of matrix operations involved in so doing and rows and columns of W_o may require transposition when any of the variance component estimates are near zero. The algorithm developed in the next section consists, in essence, of a single matrix operation and never requires row or column transformations of W_o while using the same amount of storage as the algorithm presented in [4].

2. Algorithm Developed

As described in [3], the algebraic definition of the W transformation is

$$W = W_o - L'[V'V + D^{-1}]^{-1}L \tag{5}$$

where $L = [V'V|V'X|V'Y]$ and D is the mxm diagonal matrix

$$D = \begin{bmatrix} \gamma_1 I_{m_1} & & & & \\ & \gamma_2 I_{m_2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & \gamma_c I_{m_c} \end{bmatrix}$$

Suppose that we now consider the matrix

$$\left[\begin{array}{c|c} V'V + D^{-1} & L \\ \hline L' & W_o \end{array}\right] \tag{6}$$

It is well known (see for example Goodnight [2]) that any of the sequential, in place elimination methods--Gauss, Gauss-Jordan, sequential Doolittle, sequential Cholesky--applied to the entire matrix (6) by pivoting (sweeping) on each diagonal of $V'V+D^{-1}$ will overwrite or replace W_o with W. For all of these procedures applied to a symmetric matrix, an element a_{ij} of the matrix, prior to pivoting on the (r,r) element, which is below the rth row and to the right of the rth column is replaced algebraically with

$$a_{ij} - a_{ri}a_{rj}/a_{rr} \ . \qquad (7)$$

The manner of computing (7) varies with the procedure used; however, in all cases only one multiplication is necessary. We will consider the Gauss or sequential Doolittle procedure in which we pivot sequentially on the diagonal elements of $V'V+D^{-1}$ in (6). In so doing, all elements in the rth (pivot) row or in those rows above the rth row are left unchanged when we pivot on the (r,r) element.

Obviously, the creation of the matrix (6) would greatly increase the amount of computer storage needed to perform the transformation and defeat the purpose of a compact algorithm; however, with the exception of D^{-1}, all of the elements needed to compute (7) for the elements in the lower right-hand corner of (6) are contained within that submatrix itself. Consequently, there is no reason to construct or store the matrix (6).

Let us denote the initial elements of W_o by $a_{ij}^{(o)}$ and the diagonal elements of $V'V+D^{-1}$ by $d_\ell^{(o)}$ so that as a result of the rth pivot operation, these elements would become $a_{ij}^{(r)}$ and $d_\ell^{(r)}$ respectively. Then it is easily seen that, corresponding to (7),

$$a_{ij}^{(r)} = a_{ij}^{(r-1)} - a_{ri}^{(r-1)} \cdot a_{rj}^{(r-1)} / d_r^{(r-1)} \quad i \leq j \tag{8}$$

and

$$d_\ell^{(r)} = d_\ell^{(r-1)} - \left\{ a_{r\ell}^{(r-1)} \right\}^2 / d_r^{(r-1)} \quad \ell > r. \tag{9}$$

A further simplification results by substituting $i = j = \ell$ into (8) and then subtracting this equation from (9) to obtain

$$d_\ell^{(r)} - a_{\ell\ell}^{(r)} = d_\ell^{(r-1)} - a_{\ell\ell}^{(r-1)} \quad \ell > r. \tag{10}$$

Then

$$d_\ell^{(1)} - a_{\ell\ell}^{(1)} = d_\ell^{(0)} - a_{\ell\ell}^{(0)} = (D^{-1})_{\ell\ell} \quad \ell > 1. \tag{11}$$

and, proceeding inductively, we determine that

$$d_\ell^{(r)} - a_{\ell\ell}^{(r)} = (D^{-1})_{\ell\ell} \quad \ell > r. \tag{12}$$

Thus we may write

$$d_r^{(r-1)} = a_{rr}^{(r-1)} + (D^{-1})_{rr} \tag{13}$$

and omit the storage and updating of the elements $d_\ell^{(r)}$ as suggested by (9).

The simplified algorithm is then described by (8) and (13). In order to properly apply equation (8) for the rth pivotal step it is necessary to store the elements $a_{ir}^{(r-1)}$, $i < r$ and $a_{ri}^{(r-1)}$, $i \geq r$ in a temporary work vector. The elements of D will be needed, and updated, by the optimization step and are part of the overall storage requirements.

3. Other Considerations

The W transformation algorithm in [4] included an extension to handle zero or near zero components ($\hat{\gamma}_i \doteq 0$); however, a potential re-ordering of some of the rows and columns of W_o was part of the logic of the algorithm in dealing with this situation. The simplified algorithm described in the previous section will produce the same results as the algorithm in [4] without any re-ordering of W_o. This is accomplished merely by not pivoting on any diagonal element (13) associated with a near zero component.

When none of the components $\hat{\gamma}_i$ are zero or near zero, the determinant of the matrix $V'V+D^{-1}$ is needed for subsequent evaluation of the likelihood function and is a by-product of the W transformation (see [3]). We refer the reader to [4] for the form of this matrix when zero components are involved and note that the proper log of the determinant is calculated with the simplified algorithm by taking the sum of the log of the elements (13) which are used as pivotal elements.

4. The Algorithm

Notation:

- M = Number of Random Columns in W_o
- P = Total number of columns in W_o
- W = The PxP matrix containing W_o elements in the upper triangular portion - a two dimensional matrix is used here for algorithmic simplicity.
- A = a work vector of length P
- D = vector of length M containing the diagonal elements of D
- EPSILON= the smallest value of a $\hat{\gamma}_i$ which will be deemed non-zero (usually .001 or smaller).

The following PL/1 instructions form the nucleous of the simplified W-transformation. Declaration of the necessary arrays and variables should be obvious and are omitted.

```
LOGDET = 0;                                        (A1)
DO  K = 1 TO M;                                    (A2)
   IF D(K) > EPSILON THEN DO;                      (A3)
      S = W(K,K)+1/D(K);                           (A4)
      LOGDET = LOGDET + LOG(S);                    (A5)
      DO  J = 1 TO P;                              (A6)
         IF K < J THEN A(J) = W(K,J);              (A7)
         ELSE A(J) = W(J,K); END;                  (A8)
      DO  I = 1 TO P; B = A(I)/S;                  (A9)
         DO  J = I TO P;                           (A10)
            W(I,J) = W(I,J) - B*A(J); END;         (A11)
         END; END; END;                            (A12)
```

In the above algorithm statement, (A1) initializes the log of the determinant to zero. Statement (A2) advances the index K over the range of the D vector. Statement (A3) checks the value of the K^{th} element of D and only if it is greater than EPSILON does processing continue for the current K index value. The pivotal element given by (13) is computed in statement (A4). In statement (A5) the log of the determinant is updated. Statements (A6-A8) extract the complete K^{th} row of W and store it in the A vector and statements (A9-A11) make the adjustments given by (8) to the W matrix. If a Cholesky algorithm is preferred, each element of A could be divided by the square root of S as it is stored. The only other change necessary would be to let B = A(I) in (A9) instead of B = A(I)/S.

5. Summary

The W transformation algorithm described here employs approximately the same number of multiplications/divisions as does the algorithm described in [4]. However, its simplicity reduces other overhead operations involving possible subroutine linkage and subscripting. Its compactness makes it easily codable as an assembly language routine.

6. Acknowledgments

The research on this paper by W. J. Hemmerle was supported in part by NSF grant MCS74-15331 A01.

References

[1] Corbeil, R. R. and Searle, S. R. (1976). Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. Technometrics, 18, 31-38.

[2] Goodnight, J. H. (1978). The Sweep Operator: Its Importance in Statistical Computing. Proceedings of the Computer Science and Statistics: Eleventh Annual Symposium on the Interface. Institute of Statistics, N. C. State Univ., Raleigh, N. C.

[3] Hemmerle, W. J., and Hartley, H. O. (1973). Computing Maximum Likelihood Estimates for the Mixed A.O.V. Model Using the W Transformation. Technometrics, 15, 819-831.

[4] Hemmerle, W. J., and Lorens, J. A. (1976). Improved Algorithm for the W-Transform in Variance Component Estimation. Technometrics, 18, 207-211.

[5] Hemmerle, W. J. and Downs, B. W. (1978). Nonhomogeneous Variances in the Mixed AOV Model; Maximum Likelihood Estimation. Contributions to Survey Sampling and Applied Statistics--Papers in Honor of H. O. Hartley. H. A. David, ed., N. Y., Academic Press, 153-172.

[6] Jennrich, R. I. and Sampson, P. F. (1976). Newton-Raphson and Related Algorithms for Maximum Likelihood Variance Component Estimation. Technometrics, 18, 11-18.

[7] Jennrich, R. I., and Sampson, P. F. (1978). Some Problems Faced in Making a Variance Component Algorithm into a General Mixed Model Program. Proceedings of the Computer Science and Statistics: Eleventh Annual Symposium on the Interface. Institute of Statistics, N. C. State Univ., Raleigh, N. C.

[8] Liu, L. and Sentura, J. (1977). Computation of MINQUE Variance Component Estimates. J. Amer. Statist. Assoc., 72, 867-869.

[9] Thompson R. (1975). A Note on the W Transformation. Technometrics, 17, 511-512.

Index

L

M

R

V

W