



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information**

**Citation for published version:**

JOHNSON, DL & THOMPSON, R 1995, 'Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information' Journal of Dairy Science, vol 78, no. 2, pp. 449-456.

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher final version (usually the publisher pdf)

**Published In:**

Journal of Dairy Science

**Publisher Rights Statement:**

© 1995 American Dairy Science Association

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Restricted Maximum Likelihood Estimation of Variance Components for Univariate Animal Models Using Sparse Matrix Techniques and Average Information

D. L. JOHNSON<sup>1</sup>

AgResearch  
Ruakura Agricultural Centre  
Private Bag 3123  
Hamilton, New Zealand

R. THOMPSON

Agricultural and Food Research Council  
Roslin Institute  
Roslin, Midlothian EH25 9PS, Scotland

## ABSTRACT

An algorithm is described to estimate variance components for a univariate animal model using REML. Sparse matrix techniques are employed to calculate those elements of the inverse of the coefficient matrix required for the first derivatives of the likelihood. Residuals and fitted values for random effects can be used to derive additional right-hand sides for which the mixed model equations can be repeatedly solved in turn to yield an average of the observed and expected second derivatives of the likelihood function.

This Newton method, using average information, generally converges in <10 iterations. Although the time required per iteration is two to three times greater than that required per likelihood evaluation for derivative-free methods, the total time to convergence is generally much less. An example of a complex model, involving correlated direct and maternal genetic effects, and an additional uncorrelated random effect, indicates that REML, using average information, is about five times faster than a derivative-free algorithm, using the simplex method, which is about three times faster than an expectation-maximization algorithm.

(Key words: restricted maximum likelihood, animal model, average information, sparse matrix)

Abbreviation key: AI = average information, DF = derivative-free, EM = expectation-maximization, MME = mixed model equations.

## INTRODUCTION

The REML method is preferred for estimation of variance components for animal breeding applications. For animal models, the derivative-free (DF) method has become popular because of its computational feasibility; DF only requires the calculation of the determinant of each iteration and uses search techniques to locate the maximum of the likelihood (5, 10). However, the DF method has poorer numerical properties; the solution has about one-half the accuracy, in significant digits, of the likelihood function being maximized [(14); p. 277]. This problem is likely to be greater when several variance components are estimated.

The first derivative of the likelihood has terms involving the inverse of the coefficient matrix and therefore has been regarded as expensive to calculate for large models. Misztal and Perez-Enciso (12) described the method of Takahashi et al. (16) for calculating the sparse inverse of the coefficient matrix, which calculates only those elements of the inverse that belong to the sparse pattern of the original matrix. Cost of computer time for calculation of the sparse inverse is two to three times more expensive than for calculation of the determinant, which enhances the computational

---

Received April 18, 1994.

Accepted August 26, 1994.

<sup>1</sup>New address: Livestock Improvement Corporation, Private Bag 3016, Hamilton, New Zealand.

feasibility of an expectation-maximization (EM) algorithm.

Newton's methods use first and second derivatives of the likelihood and are generally more rapid to converge. The second derivative, as used in the Newton-Raphson method, and its expected value, as used in Fisher's method of scoring, include terms involving traces of products of the inverse matrix, which represents another order of computational complexity. However, when the observed and expected second derivatives are averaged, the trace term is cancelled out, and the remaining expression is simple to compute. The objective of this paper is to discuss the use of average information (AI) REML and to compare this method with DF and EM algorithms using an example.

## MATERIALS AND METHODS

### Model

We consider the linear model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad [1]$$

where

- $\mathbf{y}$  = a vector of  $N$  observations,
- $\mathbf{b}$  = a vector of  $N_f$  fixed effects including covariables,
- $\mathbf{X}$  = the  $N \times N_f$  design matrix for fixed effects,
- $\mathbf{u}$  = a vector of  $N_u$  random effects,
- $\mathbf{Z}$  = the  $N \times N_u$  design matrix for random effects, and
- $\mathbf{e}$  = the vector of  $N$  residuals.

For the (co)variance structure of  $\mathbf{y}$  the assumption is

$$\text{var}(\mathbf{u}) = \mathbf{G},$$

$$\text{var}(\mathbf{e}) = \mathbf{R},$$

and

$$\text{cov}(\mathbf{u}, \mathbf{e}) = \mathbf{0},$$

which gives

$$\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}. \quad [2]$$

Without loss of generality,  $\mathbf{X}$  is assumed to be of full-column rank  $r(\mathbf{X}) = N_f$ .

### The Likelihood

For  $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \mathbf{V})$ , the likelihood function,  $l$ , based on linear functions of the observation vector with expectation zero, can be expressed as [e.g., (6)]

$$L = -2\ln l = \text{constant} + \ln|\mathbf{V}| + \ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y} \quad [3]$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}.$$

Using results on matrix differentiation [e.g., (15), Appendix M.7.], the first and second derivatives of the log likelihood function with respect to scalar variance components  $\theta$  and  $\phi$  can be written [e.g., (6)]

$$\frac{\partial L}{\partial \theta} = \text{tr}\left(\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{P}\right) - \mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{P}\mathbf{y} \quad [4]$$

and

$$\frac{\partial^2 L}{\partial \theta \partial \phi} = -\text{tr}\left(\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{P}\right) + 2\mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{P}\mathbf{y}. \quad [5]$$

The transition from Equations [4] to [5] assumes that  $\mathbf{V}$  is a linear function of variance components. The expected value of the second derivative is then

$$E\left(\frac{\partial^2 L}{\partial \theta \partial \phi}\right) = \text{tr}\left(\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{P}\right). \quad [6]$$

Addition of Equations [5] and [6] yields an expression for the average of observed and expected information with respect to  $\theta$  and  $\phi$ :

$$\begin{aligned} \text{AI}(\theta, \phi) &= \frac{1}{2}\left[\frac{\partial^2 L}{\partial \theta \partial \phi} + E\left(\frac{\partial^2 L}{\partial \theta \partial \phi}\right)\right] \\ &= \mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{P}\mathbf{y} \end{aligned} \quad [7]$$

and definition of the vector  $\mathbf{f}(\theta) = \frac{\partial \mathbf{V}}{\partial \theta}\mathbf{P}\mathbf{y}$  allows reexpression of Equation [7] as

$$AI(\theta, \phi) = \mathbf{f}(\theta)' \mathbf{P} \mathbf{f}(\phi). \quad [8]$$

Because  $\mathbf{P}$  is the matrix transforming observations into residuals (Appendix 1), the  $\mathbf{P} \mathbf{f}(\theta)$  are the residuals from solving the mixed model equations (MME) (Appendix 1, Equation [A2]) with  $\mathbf{y}$  replaced by  $\mathbf{f}(\theta)$ , and  $AI(\theta, \phi)$  can be computed simply as a vector product. The form of the vector  $\mathbf{f}(\theta)$  is described in the specific models.

For a univariate model, an alternative parameterization in terms of variance ratios can be achieved by factoring the residual variance out of the variance matrix  $\mathbf{V}$ , effectively reducing the dimensions of the problem by unity (1). The form of the likelihood function and its derivatives is presented in Appendix 2.

The AI REML algorithm avoids the evaluation of traces of large matrices in Equations [5] and [6] by using the AI, Equation [8], in place of second derivatives in a Newtonian-type procedure.

Let  $\Theta$  be the vector of variance components to be estimated for a particular model. Let  $AI(\Theta)$  denote the matrix with elements  $AI(\theta, \phi)$  for  $\theta, \phi \in \Theta$ , and  $DL(\Theta)$  the vector with elements  $\frac{\partial L}{\partial \theta}$  for  $\theta \in \Theta$ .

**Specific Models**

*Individual Animal Model.* Consider a univariate model with identically and independently distributed errors,  $\mathbf{R} = \mathbf{I}\sigma^2$ , one random effect  $\mathbf{u} = \mathbf{a}$ , design matrix  $\mathbf{Z} = \mathbf{Z}_a$ , and variance matrix  $\mathbf{G} = \mathbf{A}\sigma_a^2$ , where  $\mathbf{A}$  is the relationship matrix. Let  $\mathbf{C}$  denote the coefficient matrix of the MME (Appendix 1). Then, from Equation [4], the first derivatives of the log likelihood (using Equations [A4] and [A5], Appendix 1) are

$$\begin{aligned} \frac{\partial L}{\partial \sigma_a^2} &= \text{tr}(\mathbf{A} \mathbf{Z}'_a \mathbf{P} \mathbf{Z}_a) - \mathbf{y}' \mathbf{P} \mathbf{Z}_a \mathbf{A} \mathbf{Z}'_a \mathbf{P} \mathbf{y} \\ &= \frac{N_a}{\sigma_a^2} - \frac{\text{tr}(\mathbf{A}^{-1} \mathbf{C}^{aa})}{\sigma_a^4} - \frac{\hat{\mathbf{a}}' \mathbf{A}^{-1} \hat{\mathbf{a}}}{\sigma_a^4} \end{aligned} \quad [9a]$$

where  $N_a$  is the number of animals, and  $\mathbf{C}^{aa}$  is the partition of  $\mathbf{C}^{-1}$  corresponding to  $\mathbf{a}$ . Similarly,

$$\begin{aligned} \frac{\partial L}{\partial \sigma^2} &= \text{tr}(\mathbf{P}) - \mathbf{y}' \mathbf{P}^2 \mathbf{y} \\ &= \frac{N - r(\mathbf{X})}{\sigma^2} - \frac{1}{\sigma^2} \left[ N_a - \frac{\text{tr}(\mathbf{A}^{-1} \mathbf{C}^{aa})}{\sigma_a^2} \right] - \frac{\hat{\mathbf{e}}' \hat{\mathbf{e}}}{\sigma^4} \end{aligned} \quad [9b]$$

where  $\hat{\mathbf{e}}$  is the vector of residuals defined in Appendix 1. The vectors required for the AI matrix are

$$\mathbf{f}(\sigma_a^2) = \mathbf{Z}_a \mathbf{A} \mathbf{Z}'_a \mathbf{P} \mathbf{y} = \frac{1}{\sigma_a^2} \mathbf{Z}_a \hat{\mathbf{a}}, \quad [9c]$$

a scalar multiple of the vector of fitted values for random effects, and

$$\mathbf{f}(\sigma^2) = \mathbf{P} \mathbf{y} = \frac{1}{\sigma^2} \hat{\mathbf{e}}, \quad [9d]$$

a scalar multiple of the vector of residuals.

*Repeated Records and Litter Effects.* The individual animal model was extended to include an additional uncorrelated random effect  $\mathbf{c}$ , such as with repeated records on each animal, or a litter effect to consider the common environmental variance for full siblings. The random vector  $\mathbf{u}$  now represents a linking of the animal effect  $\mathbf{a}$  and the additional random effect  $\mathbf{c}$  of length  $N_c$ ,  $\mathbf{u}' = [\mathbf{a}' \mathbf{c}']$ , with incidence matrix  $\mathbf{Z} = [\mathbf{Z}_a | \mathbf{Z}_c]$  and variance matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_c^2 \end{bmatrix}.$$

In addition to Equations [9]

$$\begin{aligned} \frac{\partial L}{\partial \sigma_c^2} &= \text{tr}(\mathbf{Z}'_c \mathbf{P} \mathbf{Z}_c) - \mathbf{y}' \mathbf{P} \mathbf{Z}_c \mathbf{Z}'_c \mathbf{P} \mathbf{y} \\ &= \frac{N_c}{\sigma_c^2} - \frac{\text{tr}(\mathbf{C}^{cc})}{\sigma_c^4} - \frac{\hat{\mathbf{c}}' \hat{\mathbf{c}}}{\sigma_c^4} \end{aligned} \quad [10a]$$

where  $\mathbf{C}^{cc}$  is the partition of  $\mathbf{C}^{-1}$  corresponding to  $\mathbf{c}$ , and

$$\mathbf{f}(\sigma_c^2) = \mathbf{Z}_c \mathbf{Z}'_c \mathbf{P} \mathbf{y} = \frac{1}{\sigma_c^2} \mathbf{Z}_c \hat{\mathbf{c}}. \quad [10b]$$

*Maternal Effects.* The model was further generalized to include a second animal effect

$\mathbf{m}$  such that  $\mathbf{u}' = [\mathbf{a}'\mathbf{m}'\mathbf{c}']$ ,  $\mathbf{Z} = [\mathbf{Z}_a|\mathbf{Z}_m|\mathbf{Z}_c]$ , and

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{am} & \mathbf{0} \\ \mathbf{A}\sigma_{am} & \mathbf{A}\sigma_m^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_c^2 \end{bmatrix}$$

Let  $\mathbf{G}_0 = \begin{bmatrix} \sigma_a^2 & \sigma_{am} \\ \sigma_{am} & \sigma_m^2 \end{bmatrix}$

with inverse  $\mathbf{G}_0^{-1} = \begin{bmatrix} \sigma^{aa} & \sigma^{am} \\ \sigma^{am} & \sigma^{mm} \end{bmatrix}$ .

Then, instead of [9a] and [9c], the first derivatives are

$$\begin{bmatrix} \frac{\partial L}{\partial \sigma_a^2} & \frac{1}{2} \frac{\partial L}{\partial \sigma_{am}} \\ \frac{1}{2} \frac{\partial L}{\partial \sigma_{am}} & \frac{\partial L}{\partial \sigma_m^2} \end{bmatrix} = \mathbf{N}_a \mathbf{G}_0^{-1} - \mathbf{G}_0^{-1} \begin{bmatrix} \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{aa}) + \hat{\mathbf{a}}' \mathbf{A}^{-1} \hat{\mathbf{a}} & \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{am}) + \hat{\mathbf{a}}' \mathbf{A}^{-1} \hat{\mathbf{m}} \\ \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{am}) + \hat{\mathbf{a}}' \mathbf{A}^{-1} \hat{\mathbf{m}} & \text{tr}(\mathbf{A}^{-1}\mathbf{C}^{mm}) + \hat{\mathbf{m}}' \mathbf{A}^{-1} \hat{\mathbf{m}} \end{bmatrix} \mathbf{G}_0^{-1} \quad [11a]$$

where  $\mathbf{C}^{aa}$ ,  $\mathbf{C}^{am}$ , and  $\mathbf{C}^{mm}$  are partitions of  $\mathbf{C}^{-1}$ , and

$$\begin{aligned} \mathbf{f}(\sigma_a^2) &= \mathbf{Z}_a(\sigma^{aa}\hat{\mathbf{a}} + \sigma^{am}\hat{\mathbf{m}}) \\ \mathbf{f}(\sigma_{am}) &= \mathbf{Z}_a(\sigma^{am}\hat{\mathbf{a}} + \sigma^{mm}\hat{\mathbf{m}}) \\ &\quad + \mathbf{Z}_m(\sigma^{aa}\hat{\mathbf{a}} + \sigma^{am}\hat{\mathbf{m}}) \\ \mathbf{f}(\sigma_m^2) &= \mathbf{Z}_m(\sigma^{am}\hat{\mathbf{a}} + \sigma^{mm}\hat{\mathbf{m}}). \end{aligned} \quad [11b]$$

**Computational Procedure**

The sparse matrix package FSPAK (Perez-Enciso and Misztal, 1993, personal communication) was used for the calculations. The package is an interface to code for sparse matrix methods by George and Liu (4) and provided solutions to the MME, the sparse matrix inverse and log determinant required in the following.

*Step 1.* Pedigree and data files were read, and the inverse of the relationship matrix was calculated.

*Step 2.* The MME were ordered using the minimum degree ordering subroutines of Ge-

orge and Liu (4) to minimize the additional nonzero elements generated on factorization.

*Step 3.* Symbolic factorization was carried out to determine the data structure of the upper triangular factor.

*Step 4.* Numerical values were input to the MME, given current estimates of the variance components, and numerical factorization of the coefficient matrix was carried out. The equations were solved, and residuals were calculated.

*Step 5.* The elements of the average information matrix were calculated. Given a variance component  $\theta \in \Theta$ , the vector  $\mathbf{f}(\theta)$  was calculated based on fitted values or residuals as shown in Equations [9c,d], [10b], and [11b]. This vector was used as a right-hand side, and, with the same coefficient matrix set up at step 4, the MME were solved. The residuals  $\mathbf{P}\mathbf{f}(\theta)$  and the vector products  $\mathbf{f}(\phi)\mathbf{P}\mathbf{f}(\theta)$  were calculated for all  $\phi \in \Theta$ . Step 5 was repeated for all  $\theta \in \Theta$ .

*Step 6.* The sparse inverse of the coefficient matrix was evaluated using the method of Takahashi et al. (16). The trace terms and vector products required for Equations [9a, 9b], [10a], and [11a] were calculated to yield the first derivatives of the likelihood.

*Step 7.* The variance components were updated using Newton's method:

$$\Theta^{[p+1]} = \Theta^{[p]} - [\mathbf{AI}(\Theta^{[p]})]^{-1}[\mathbf{DL}(\Theta^{[p]})] \quad [12]$$

where [p] is iterate number.

Steps 4 to 7 were repeated until convergence. The inverse of the information matrix was used to estimate the (co)variances of the variance components. Calculation of the sparse inverse in step 6 is the most computationally demanding; the repeated solving in step 5 is relatively inexpensive.

**Example**

The example involves 2307 records of weaning weight from a flock of 2500 Romney sheep over 15 yr. Direct and maternal genetic effects and a permanent environmental effect to take account of the additional nongenetic correlation between progeny from the same dam (695 dams) were fitted. Fixed effects included 30 contemporary groups for interaction

of year and sex, three classes for birth and rearing rank, four classes for age of dam, and a covariant for date of birth. The rank of the coefficient matrix of the MME was 5731 with 53,509 nonzeros.

The variance ratio parameterization was used (see Appendix 2). Two sets of starting values were considered for  $\{\sigma_a^2, \sigma_m^2, \sigma_{am}, \sigma_c^2\}$  the direct, maternal, direct and maternal, and permanent environmental (co)variances, respectively. Expressed as a proportion of the phenotypic variance, the two sets were  $\{.3, .1, -.1, .2\}$  and  $\{.1, .1, .02, .1\}$ . The convergence criterion for AI REML was  $\|\Delta\lambda\|/\|\lambda\| < 10^{-3}$  where  $\|\cdot\|$  denotes vector length (Euclidean norm), and  $\lambda$  is the vector of variance ratios.

For a comparison with a DF algorithm, the simplex method (13) was used in place of steps 5 to 7. For the EM algorithm, steps 5 and 7 were not required, and the variance parameter estimates were updated using an algorithm similar to that used by Cantet et al. (2) for a similar model. For example, the variance parameter  $\sigma_a^2$  was updated using

$$\sigma_a^{2[p+1]} = \frac{1}{(1 - \beta)} \left[ \frac{\hat{a}' A^{-1} \hat{a}}{N_a - \text{tr}(A^{-1} C^{aa})} \sigma_a^{2[p]} - \beta \sigma_a^{2[p]} \right]$$

where  $\beta$  was chosen as .5, with similar expressions for the other variance components.

The DF and EM procedures were stopped when their likelihood values attained the maximum value achieved by AI REML to a specified degree of accuracy. As verification of the equivalence of convergence among the three algorithms, the final variance estimates from the DF and EM algorithms, when used as starting values for AI REML, were required to satisfy the convergence criterion for AI REML without further iteration.

**RESULTS AND DISCUSSION**

The log likelihood values agreed within five decimal places for the three algorithms. The final parameter estimates ( $\pm$ SE) were  $\{.197 \pm .039, .117 \pm .037, .004 \pm .027, \text{ and } .119 \pm .027\}$ , and residual variance was 5.217. The standard errors were derived from the inverse of the AI matrix.

TABLE 1. Number of iterations, number of likelihood evaluations, and computing time for the average information (AI), derivative-free (DF), and expectation-maximization (EM) REML algorithms for the weaning weight example.

	AI	DF	EM
Starting values {.3, .1, -.1, .2}			
Iterations, no.	6		109
Likelihoods, no.		169	
Computer time, s	234	1238	4106
Starting values {.1, .1, .02, .1}			
Iterations, no.	5		64
Likelihoods, no.		119	
Computer time, s	192	872	2403

Table 1 summarizes the computer time and number of iterations for the three algorithms (Compaq Deskpro 4/33i; Compaq Computer Corporation, Houston, TX) using the Lahey F77L-EM/32 Fortran compiler (Lahey Computer Systems Inc., Incline Village, NV). The DF algorithm shows the number of likelihood evaluations, because several evaluations are carried out at each iteration, and likelihood evaluation is the major cost for this method. The AI REML procedure was about five times faster than DF, which was about three times faster than EM. These results should be viewed as indicative of the relative performance of the three algorithms, because speed rankings can vary among computing platforms. Furthermore, the rankings are likely to depend on starting values, the initial simplex chosen for the DF method, and the EM updating procedure used. Also, DF methods other than the simplex method have not been considered here.

An acceleration procedure was used with the EM algorithm. The second method of Laird et al. (7), which effectively attempts to approximate the matrix of second derivatives, was unsuccessful, as also found by Misztal (11). An exponential (Aitken) extrapolation, the first method of Laird et al. (7) as derived from the average of the ratios of the differences of the individual parameter estimates obtained for the two most recent iterations, was effective initially but led to slow convergence for the latter iterations. The method employed in this example considered the ratio of differences for each

component separately, which seemed to be more effective, although the cause is not clear, and no algebraic proof of this result is offered. This method is similar to independent application of a univariate acceleration to each parameter. The acceleration was applied about every fifth iteration, provided that the likelihood function was increased and that parameters remained within boundary limits.

Possible criticisms of Newton's methods are that global convergence is not guaranteed, the Hessian matrix may not be positive definite, or parameter estimates may exceed the boundary. The AI matrix, Equation [8], is composed of residual sums of squares and cross-products and is therefore positive semi-definite. Globally convergent modifications of Newton's method can be employed if, at any iteration, a Newton step is inappropriate (3). For example, Equation [12] could be modified using Marquardt's method (9):

$$\Theta^{[p+1]} = \Theta^{[p]} - [\mathbf{AI}(\Theta^{[p]}) + \mu\mathbf{I}]^{-1}[\mathbf{DL}(\Theta^{[p]})],$$

where a constant,  $\mu$ , is added to the diagonal elements of the Hessian. This method is a compromise between Newton's method ( $\mu = 0$ ) and steepest ascent ( $\mu$  large), and the smallest value of  $\mu$  is chosen so that the step taken increases the likelihood function while the new variance component estimates are confined within their domain of definition.

The model trust region approach (3), to select a suitable value for  $\mu$ , has been used in the program code for AI REML to protect against starting values that are too distant from the maximum or for which variance components become negative or correlations exceed unity in absolute value. The approach is based on estimation of the region about the current estimates in which the local quadratic model underlying Newton's method can be trusted to represent adequately the likelihood function and to take a step to approximately maximize the likelihood function in this region.

When parameter estimates exceed the boundary, the AI REML procedure can be repeated using a simpler model if appropriate. Other parameterizations, including Cholesky transformations (8), may also be used to constrain parameters within boundary limits.

The example in this paper considered a model with several variance components. The

differences in performance among the algorithms are expected to be not as great for simpler models involving fewer variance components but greater for multivariate models.

## CONCLUSIONS

The use of a sparse matrix inverse and the calculation of second differentials from the average of observed and expected values provides an efficient computing algorithm for variance component estimation with an animal model and a very competitive alternative to DF and EM algorithms.

The use of AI is only strictly correct when parents are not selected. The more intense the selection is, the greater the discrepancy between AI and observed information; presumably, the number of iterations will increase, but DF and EM are not expected to become competitive.

Calculation of a sparse inverse increases memory requirements over that required for likelihood evaluation. Developments to reduce memory requirements (R. Thompson, N. R. Wray, and R. E. Crump, 1994, unpublished data) for finding the elements of the sparse matrix have been incorporated in FSPAK.

## ACKNOWLEDGMENTS

This work was supported by the New Zealand Foundation for Research, Science and Technology.

## REFERENCES

- 1 Callanan, T. P., and D. A. Harville. 1991. Some new algorithms for computing restricted maximum likelihood estimates of variance components. *J. Stat. Comp. Simul.* 38:239.
- 2 Cantet, R.J.C., D. Gianola, I. Misztal, and R. L. Fernando. 1993. Estimates of dispersion parameters and of genetic and environmental trends for weaning weight in Angus cattle using a maternal animal model with genetic grouping. *Livest. Prod. Sci.* 34:203.
- 3 Dennis, J. E., and R. B. Schnabel. 1983. *Numerical Methods for Unconstrained Optimization and Non-linear Equations*. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- 4 George, A., and J. W. Liu. 1981. *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, Inc., Englewood Cliffs, NJ.
- 5 Graser, H. U., S. P. Smith, and B. Tier. 1987. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.* 64:1362.

6 Harville, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72:320.  
 7 Laird, N. M., N. Lange, and D. Stram. 1987. Maximum likelihood computations with repeated measures: application of the EM algorithm. *J. Am. Stat. Assoc.* 82:97.  
 8 Lindstrom, M. J., and D. M. Bates. 1988. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *J. Am. Stat. Assoc.* 83:1014.  
 9 Marquardt, D. 1963. An algorithm for least-squares estimation of non-linear parameters. *Soc. Ind. Appl. Math. J. Appl. Math.* 11:431.  
 10 Meyer, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet. Sel. Evol.* 21:317.  
 11 Misztal, I. 1990. Restricted maximum likelihood estimation of variance components in animal model using sparse matrix inversion and a supercomputer. *J. Dairy Sci.* 73:163.  
 12 Misztal, I., and M. Perez-Enciso. 1993. Sparse matrix inversion for restricted maximum likelihood estimation of variance components by expectation-maximization. *J. Dairy Sci.* 76:1479.  
 13 Nelder, J. A., and R. Mead. 1965. A simplex method for function maximisation. *Computer J.* 7:308.  
 14 Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1989. *Numerical Recipes*. Cambridge Univ. Press, Cambridge, MA.  
 15 Searle, S. R., G. Casella, and C. E. McCulloch. 1992. *Variance Components*. John Wiley & Sons, New York, NY.  
 16 Takahashi, K., J. Fagan, and M. S. Chin. 1973. Formation of a sparse bus impedance matrix and its application to short circuit study. Page 63 in *Proc. 8th Inst. PICA Conf.*, Minneapolis, MN.

APPENDIX 1

Matrix Identities

The inverse of the variance matrix  $V$  defined in Equation [2] is given by

$$V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}, \quad [A1]$$

and the MME corresponding to the model defined by Equations [1] and [2] are

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}. \quad [A2]$$

From application of results for the inverse of a partitioned matrix [e.g., (15), Appendix M.5.] to the coefficient matrix in Equation [A2], it

follows that the matrix

$$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$$

has the equivalent form

$$P = S - SZ(Z'SZ + G^{-1})^{-1}Z'S \quad [A3]$$

where  $S$  is the absorption matrix

$$S = R^{-1} - R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}.$$

Then, using [A3],

$$Z'PZ = G^{-1} - G^{-1}(Z'SZ + G^{-1})^{-1}G^{-1}. \quad [A4]$$

Note that  $(Z'SZ + G^{-1})^{-1}$  is that partition of the inverse of the coefficient matrix corresponding to the random vector  $u$  with solution

$$\hat{u} = (Z'SZ + G^{-1})^{-1}Z'Sy.$$

From [A3],

$$Z'Py = G^{-1}\hat{u}, \quad [A5]$$

and, using [A2] and because  $SX = 0$ , then

$$Py = Sy - SZ\hat{u} = S(y - X\hat{b} - Z\hat{u}) = R^{-1}\hat{e}$$

where  $\hat{e} = y - X\hat{b} - Z\hat{u}$  is the vector of residuals. Because  $SRS = S$ , it follows from [A3]

$$\begin{aligned} \text{tr}(PR) &= \text{tr}(SR) - \text{tr}[(Z'SZ + G^{-1})^{-1}Z'SZ] \\ &= N - r(X) - N_u + \text{tr}[(Z'SZ + G^{-1})^{-1}G^{-1}]. \end{aligned} \quad [A6]$$

APPENDIX 2

Variance Ratio Parameterization

The residual variance is factored out of the variance matrix  $V$  so that  $V = V_\lambda\sigma^2$ , where  $V_\lambda = ZAZ' + I$  and  $\Lambda$  is the variance matrix  $G$  scaled by  $1/\sigma^2$ . The elements of  $\Lambda$  are variance ratios. Define the scaled matrix  $P_\lambda = P\sigma^2$ . Then the log likelihood (Equation [3]), omitting the constant, may be reexpressed as



$$L = (N - r(\mathbf{X}))\ln\sigma^2 + \ln|\mathbf{V}_\lambda| + \ln|\mathbf{X}'\mathbf{V}_\lambda^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}_\lambda\mathbf{y}/\sigma^2. \quad [\text{B1}]$$

Setting  $\partial L/\partial\sigma^2 = 0$  directly yields an estimate of the residual variance, given the vector of variance ratios ( $\lambda$ ), from the residual sum of squares

$$\hat{\sigma}^2(\lambda) = \mathbf{y}'\mathbf{P}_\lambda\mathbf{y}/(N - r(\mathbf{X})). \quad [\text{B2}]$$

Substitution of that expression in [B1] gives the concentrated likelihood function (1)

$$L_c(\lambda) = (N - r(\mathbf{X}))\ln\hat{\sigma}^2(\lambda) + \ln|\mathbf{V}_\lambda| + \ln|\mathbf{X}'\mathbf{V}_\lambda^{-1}\mathbf{X}|.$$

The concentrated likelihood attains its maximum at the REML estimates  $\hat{\lambda}$  of  $\lambda$ , in which case  $\hat{\sigma}^2(\hat{\lambda})$  is the REML estimate of  $\sigma^2$ . For the derivatives, for variance ratios  $\lambda_i$  and  $\lambda_j$ ,

$$\frac{\partial L_c}{\partial \lambda_i} = \text{tr}\left(\frac{\partial \mathbf{V}_\lambda}{\partial \lambda_i} \mathbf{P}_\lambda\right) - \mathbf{y}'\mathbf{P}_\lambda \frac{\partial \mathbf{V}_\lambda}{\partial \lambda_i} \mathbf{P}_\lambda \mathbf{y}/\hat{\sigma}^2(\lambda) \quad [\text{B3}]$$

and

$$\frac{\partial^2 L_c}{\partial \lambda_i \partial \lambda_j} = \frac{\partial^2 L}{\partial \lambda_i \partial \lambda_j} - \left(\frac{\partial^2 L}{\partial \lambda_i \partial \sigma^2}\right) \left(\frac{\partial^2 L}{\partial \lambda_j \partial \sigma^2}\right) \left(\frac{\hat{\sigma}^4(\lambda)}{N - r(\mathbf{X})}\right).$$

The derivatives on the right-hand side of [B4] are evaluated at  $\sigma^2 = \hat{\sigma}^2(\lambda)$  and have expressions similar to Equation [5]. The average of observed and expected values of these derivatives are

$$\begin{aligned} \text{AI}(\lambda_i, \lambda_j) &= \mathbf{y}'\mathbf{P}_\lambda \frac{\partial \mathbf{V}_\lambda}{\partial \lambda_i} \mathbf{P}_\lambda \frac{\partial \mathbf{V}_\lambda}{\partial \lambda_j} \mathbf{y}/\hat{\sigma}^2(\lambda) \\ \text{AI}(\lambda_i, \sigma^2) &= \left[ \text{tr}\left(\mathbf{P}_\lambda \frac{\partial \mathbf{V}_\lambda}{\partial \lambda_i}\right) + \left[ \mathbf{y}'\mathbf{P}_\lambda \frac{\partial \mathbf{V}_\lambda}{\partial \lambda_i} \mathbf{P}_\lambda \mathbf{y}/\hat{\sigma}^2(\lambda) \right] \right] / 2\hat{\sigma}^2(\lambda), \quad [\text{B5}] \end{aligned}$$

and these can be substituted into the right-hand side of [B4] in place of the corresponding observed values to yield average information for the concentrated likelihood.