

# DERIVATIVE-INTENSE RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION OF COVARIANCE COMPONENTS FOR ANIMAL MODELS

K. Meyer

Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia

## SUMMARY

The calculation of derivatives of the likelihood function for animal models without the need for large matrix inversion is described. Their use in estimating covariance components is illustrated giving examples of analyses of beef cattle data.

## INTRODUCTION

Estimation of covariance components by Restricted Maximum Likelihood (REML) fitting an animal model is widely carried out using a derivative-free (DF) algorithm to locate the maximum of the likelihood function. While such algorithms have proven robust and easy to use, they are generally slow to converge, often requiring many likelihood evaluations, in particular for multivariate analyses fitting several random factors. As outlined by Graser *et al.* (1987), however, they only require factorization of a large matrix rather than its inverse and can be implemented efficiently using sparse matrix techniques for analyses involving tens of thousands of animals.

Although there has been some use of algorithms using derivatives of the likelihood function, i.e. Expectation-Maximization type or even Method of Scoring procedures, for large scale animal model analyses, they have involved the use of a supercomputer or some approximation of the inverse of the coefficient matrix required (Ducrocq, 1993; Misztal, 1990; Misztal *et al.*, 1992). This paper describes the calculation of first and second derivatives of the REML (log) likelihood ( $\log L$ ) using a simple extension of the large matrix factorization required to evaluate  $\log L$  (for a DF algorithm) only, and illustrates their use in estimating covariance components using a Newton-Raphson algorithm.

## DERIVATIVES OF THE LIKELIHOOD

Consider the linear mixed model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (1)$$

with  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{u}$  and  $\mathbf{e}$  denoting the vector of observations, fixed effects, random effects and residual errors, respectively, and  $\mathbf{X}$  and  $\mathbf{Z}$  the incidence matrices pertaining to  $\mathbf{b}$  and  $\mathbf{u}$ . Let  $V(\mathbf{u}) = \mathbf{G}$ ,  $V(\mathbf{e}) = \mathbf{R}$  and  $Cov(\mathbf{u}, \mathbf{e}') = \mathbf{0}$ , so that  $V(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ . The mixed model matrix pertaining to (1) is

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{r} \\ \mathbf{r}' & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (2)$$

where  $\mathbf{C}$  is the coefficient matrix and  $\mathbf{r}$  is the vector of right hand sides in the mixed model equations.

For a multivariate normal distribution,  $\mathbf{y} \sim N(\mathbf{X}\mathbf{b}, \mathbf{V})$ ,

$$\log L = -\frac{1}{2} \left[ const + \log |\mathbf{V}| + \log |\mathbf{X}^* \mathbf{V}^{-1} \mathbf{X}^*| + (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \right] \quad (3)$$

(e.g. Harville 1977), with  $\mathbf{X}^*$  a full-rank submatrix of  $\mathbf{X}$ . Alternatively (Graser *et al.*, 1987; Meyer, 1989),

$$\log L = -\frac{1}{2} \left[ const + \log |\mathbf{R}| + \log |\mathbf{G}| + \log |\mathbf{C}| + \mathbf{y}'\mathbf{P}\mathbf{y} \right] \quad (4)$$

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

Let  $\theta = \{\theta_i\}$  denote the vector of parameters to be estimated with  $i = 1, \dots, p$ . Consider a parameterization where  $\mathbf{V}$  is linear in  $\theta$ , i.e.  $\mathbf{V} = \sum_{i=1}^p \theta_i \partial \mathbf{V} / \partial \theta_i$ . Differentiating (4) gives

$$\frac{\partial \log L}{\partial \theta_i} = -\frac{1}{2} \left[ \frac{\partial \log |\mathbf{R}|}{\partial \theta_i} + \frac{\partial \log |\mathbf{G}|}{\partial \theta_i} + \frac{\partial \log |\mathbf{C}|}{\partial \theta_i} + \frac{\partial \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial \theta_i} \right] \quad (5)$$

$$\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} = -\frac{1}{2} \left[ \frac{\partial^2 \log |\mathbf{R}|}{\partial \theta_i \partial \theta_j} + \frac{\partial^2 \log |\mathbf{G}|}{\partial \theta_i \partial \theta_j} + \frac{\partial^2 \log |\mathbf{C}|}{\partial \theta_i \partial \theta_j} + \frac{\partial^2 \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial \theta_i \partial \theta_j} \right] \quad (6)$$

### Calculating $\log |\mathbf{C}|$ and $\mathbf{y}'\mathbf{P}\mathbf{y}$ and their derivatives

Graser *et al.* (1987) showed how  $\log |\mathbf{C}|$  and  $\mathbf{y}'\mathbf{P}\mathbf{y}$  can be evaluated in a general way for all models of form (1) by factoring  $\mathbf{M}$ . An alternative to the series of Gaussian Elimination steps they suggested, is a Cholesky decomposition. Let  $\mathbf{L}$  with elements  $l_{ij}$  ( $l_{ij} = 0$  for  $j > i$ ) denote the Cholesky factor of  $\mathbf{M}$ , i.e.  $\mathbf{M} = \mathbf{L}\mathbf{L}'$ . Hence,

$$\log |\mathbf{C}| = 2 \sum_{k=1}^{M-1} \log l_{kk} \quad (7)$$

with  $M$  denoting the size of  $\mathbf{M}$ , and noting that  $\mathbf{y}'\mathbf{P}\mathbf{y} = |\mathbf{M}|/|\mathbf{C}|$  (Smith, 1993)

$$\mathbf{y}'\mathbf{P}\mathbf{y} = l_{MM}^2 \quad (8)$$

Smith (1993) describes a general algorithm which allows the derivatives of the Cholesky factor of a matrix to be evaluated while carrying out the factorization, provided the derivatives of the original matrix are specified. Differentiating (7) and (8) then gives the derivatives of  $\log |\mathbf{C}|$  and  $\mathbf{y}'\mathbf{P}\mathbf{y}$  as simple functions of the diagonal elements of the Cholesky matrix and its derivatives.

$$\frac{\partial \log |\mathbf{C}|}{\partial \theta_i} = 2 \sum_{k=1}^{M-1} l_{kk}^{-1} \frac{\partial l_{kk}}{\partial \theta_i} \quad (9)$$

$$\frac{\partial^2 \log |\mathbf{C}|}{\partial \theta_i \partial \theta_j} = 2 \sum_{k=1}^{M-1} l_{kk}^{-1} \frac{\partial^2 l_{kk}}{\partial \theta_i \partial \theta_j} - l_{kk}^{-2} \frac{\partial l_{kk}}{\partial \theta_i} \frac{\partial l_{kk}}{\partial \theta_j} \quad (10)$$

$$\frac{\partial \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial \theta_i} = 2 l_{MM} \frac{\partial l_{MM}}{\partial \theta_i} \quad (11)$$

$$\frac{\partial^2 \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial \theta_i \partial \theta_j} = 2 \left( l_{MM} \frac{\partial^2 l_{MM}}{\partial \theta_i \partial \theta_j} + \frac{\partial l_{MM}}{\partial \theta_i} \frac{\partial l_{MM}}{\partial \theta_j} \right) \quad (12)$$

### Calculating $\log |\mathbf{R}|$ and its derivatives

Consider  $q$  traits and let  $\mathbf{E}$  with elements  $e_{ij}$  ( $i \leq j = 1, \dots, q$ ) be the symmetric matrix of residual covariances. For  $\mathbf{y}$  ordered according to traits within animals and zero error covariances between measurements on different animals,  $\mathbf{R}$  is blockdiagonal with submatrices  $\mathbf{E}_w$  for the  $i$ -th animal with combination of traits  $w$ . This gives (Meyer, 1991)

$$\log |\mathbf{R}| = \sum_{w=1}^W N_w \log |\mathbf{E}_w| \quad (13)$$

where  $N_w$  is the number of animals having records for combination of traits  $w$ , and  $W$  the number of different combinations of traits measured. Analogously, derivatives of  $\log |\mathbf{R}|$  can be obtained from derivatives of matrices  $\mathbf{E}_w$ . Let  $e_w^{rs}$  denote the  $rs$ -th element of  $\mathbf{E}_w^{-1}$ , and define  $\mathbf{D}_w^{\theta_i} = \partial \mathbf{E}_w / \partial \theta_i$ . For  $\theta_i = e_{kl}$  and  $\theta_j = e_{mn}$

$$\frac{\partial \log |\mathbf{R}|}{\partial \theta_i} = \sum_{w=1}^W N_w \text{tr}(\mathbf{E}_w^{-1} \mathbf{D}_w^{\theta_i}) = \sum_{w=1}^W N_w (2 - \delta_{kl}) e_w^{kl} \quad (14)$$

$$\frac{\partial^2 \log |\mathbf{R}|}{\partial \theta_i \partial \theta_j} = \sum_{w=1}^W N_w \text{tr}(\mathbf{E}_w^{-1} \mathbf{D}_w^{\theta_i} \mathbf{E}_w^{-1} \mathbf{D}_w^{\theta_j}) = \sum_{w=1}^W N_w (2 - \delta_{kl})(2 - \delta_{mn})(e_w^{km} e_w^{ln} + e_w^{lm} e_w^{kn}) \quad (15)$$

where  $\delta_{rs}$  is Kronecker's Delta, i.e.  $\delta_{rs} = 1$  for  $r = s$  and zero otherwise, while all other derivatives of  $\log |\mathbf{R}|$  are zero. For  $q = 1$  and  $\mathbf{R} = \sigma_E^2 \mathbf{I}$ , (14) and (15) reduce to  $N\sigma_E^{-2}$  and  $2N\sigma_E^{-4}$ , respectively.

### Calculating $\log |\mathbf{G}|$ and its derivatives

Terms arising from the covariance matrix of random effects,  $\mathbf{G}$ , can often be determined in a similar way, exploiting the structure of  $\mathbf{G}$  (Meyer, 1989, 1991). Define  $\mathbf{T} = \{t_{ij}\}$  of size  $rq \times rq$  as the matrix of covariances between the  $r$  random effects fitted. Let  $\mathbf{u}$  consist of a vector of  $qN_A$  animal genetic effects,  $\mathbf{a}$ , and some uncorrelated additional random effect,  $\mathbf{c}$ , with  $N_C$  levels per trait, i.e.  $\mathbf{u}' = (\mathbf{a}'\mathbf{c}')$ . Partition  $\mathbf{T}$  correspondingly, and let  $\mathbf{A}$  denote the numerator relationship between animals, and  $\mathbf{B}$  describe the correlation structure amongst the levels of  $\mathbf{c}$ . This gives

$$\mathbf{G} = \text{Diag} \{ \mathbf{A} \times \mathbf{T}_A; \mathbf{B} \times \mathbf{T}_C \} \quad (16)$$

$$\log |\mathbf{G}| = N_A \log |\mathbf{T}_A| + N_C \log |\mathbf{T}_C| + q(\log |\mathbf{A}| + \log |\mathbf{B}|) \quad (17)$$

$$\frac{\partial \log |\mathbf{G}|}{\partial \theta_i} = N_A \text{tr}(\mathbf{T}_A^{-1} \mathbf{D}_A^{\theta_i}) + N_C \text{tr}(\mathbf{T}_C^{-1} \mathbf{D}_C^{\theta_i}) \quad (18)$$

$$\frac{\partial^2 \log |\mathbf{G}|}{\partial \theta_i \partial \theta_j} = N_A \text{tr}(\mathbf{T}_A^{-1} \mathbf{D}_A^{\theta_i} \mathbf{T}_A^{-1} \mathbf{D}_A^{\theta_j}) + N_C \text{tr}(\mathbf{T}_C^{-1} \mathbf{D}_C^{\theta_i} \mathbf{T}_C^{-1} \mathbf{D}_C^{\theta_j}) \quad (19)$$

with  $\mathbf{D}_A^{\theta_i} = \partial \mathbf{T}_A / \partial \theta_i$  and  $\mathbf{D}_C^{\theta_i} = \partial \mathbf{T}_C / \partial \theta_i$ . Further simplifications analogous to (14) and (15) can be derived; e.g. for  $\mathbf{G} = \sigma_A^2 \mathbf{A}$ , (18) and (19) reduce to  $N_A \sigma_A^{-2}$  and  $2N_A \sigma_A^{-4}$ , respectively.

### Derivatives of the mixed model matrix

Fortunately, derivatives of  $\mathbf{M}$  have the same structure as  $\mathbf{M}$  and can be evaluated while setting up  $\mathbf{M}$ , replacing  $\mathbf{G}$  and  $\mathbf{R}$  by their derivatives. For  $\theta_i$  and  $\theta_j$  equal to residual (co)variances, the derivatives of  $\mathbf{M}$  are of form

$$\begin{bmatrix} \mathbf{X}'\mathbf{Q}_R\mathbf{X} & \mathbf{X}'\mathbf{Q}_R\mathbf{Z} & \mathbf{X}'\mathbf{Q}_R\mathbf{y} \\ \mathbf{Z}'\mathbf{Q}_R\mathbf{X} & \mathbf{Z}'\mathbf{Q}_R\mathbf{Z} & \mathbf{Z}'\mathbf{Q}_R\mathbf{y} \\ \mathbf{y}'\mathbf{Q}_R\mathbf{X} & \mathbf{y}'\mathbf{Q}_R\mathbf{Z} & \mathbf{y}'\mathbf{Q}_R\mathbf{y} \end{bmatrix} \quad (20)$$

with

$$\mathbf{Q}_{R1} = - \sum_{k=1}^N + \mathbf{E}_{w_k}^{-1} \mathbf{D}_{w_k}^{\theta_i} \mathbf{E}_{w_k}^{-1} \quad (21)$$

and

$$\mathbf{Q}_{R2} = \sum_{k=1}^N + \mathbf{E}_{w_k}^{-1} \mathbf{D}_{w_k}^{\theta_i} \mathbf{E}_{w_k}^{-1} \mathbf{D}_{w_k}^{\theta_j} \mathbf{E}_{w_k}^{-1} + \mathbf{E}_{w_k}^{-1} \mathbf{D}_{w_k}^{\theta_j} \mathbf{E}_{w_k}^{-1} \mathbf{D}_{w_k}^{\theta_i} \mathbf{E}_{w_k}^{-1} \quad (22)$$

for first and second derivatives, respectively, and  $w_k$  denoting the combination of traits for the  $k$ -th animal and  $\Sigma^+$  the direct matrix sum. For  $\mathbf{R} = \sigma_E^2 \mathbf{I}$  and  $\theta_i = \theta_j = \sigma_E^2$ ,  $\mathbf{Q}_{R1} = -\sigma_E^{-4} \mathbf{I}$  and  $\mathbf{Q}_{R2} = \sigma_E^{-6} \mathbf{I}$ . Analogously, for  $\theta_i$  and  $\theta_j$  equal to elements of  $\mathbf{T}$  and  $\mathbf{G}$  as in (16), derivatives of  $\mathbf{M}$  are of form

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_G & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (23)$$

with

$$\mathbf{Q}_{G1} = \begin{bmatrix} -\mathbf{T}_A^{-1} \mathbf{D}_A^{\theta_i} \mathbf{T}_A^{-1} \times \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{T}_C^{-1} \mathbf{D}_C^{\theta_i} \mathbf{T}_C^{-1} \times \mathbf{B}^{-1} \end{bmatrix} \quad (24)$$

and  $\mathbf{Q}_{G2} =$

$$\begin{bmatrix} \mathbf{T}_A^{-1} (\mathbf{D}_A^{\theta_i} \mathbf{T}_A^{-1} \mathbf{D}_A^{\theta_j} + \mathbf{D}_A^{\theta_j} \mathbf{T}_A^{-1} \mathbf{D}_A^{\theta_i}) \mathbf{T}_A^{-1} \times \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_C^{-1} (\mathbf{D}_C^{\theta_i} \mathbf{T}_C^{-1} \mathbf{D}_C^{\theta_j} + \mathbf{D}_C^{\theta_j} \mathbf{T}_C^{-1} \mathbf{D}_C^{\theta_i}) \mathbf{T}_C^{-1} \times \mathbf{B}^{-1} \end{bmatrix} \quad (25)$$

Again, further simplifications are feasible for specific cases. For  $q = 1$ ,  $\mathbf{G} = \sigma_A^2 \mathbf{A}$  and  $\theta_i = \theta_j = \sigma_A^2$ ,  $\mathbf{Q}_{G1} = -\sigma_A^{-4} \mathbf{A}^{-1}$  and  $\mathbf{Q}_{G2} = \sigma_A^{-6} \mathbf{A}^{-1}$ .

## IMPLEMENTATION

Variance components were estimated using a Newton-Raphson (NR) type algorithm. The NR is an unconstrained optimization procedure. REML, however, requires estimates to be within the permissible parameter space. Standard procedures to impose constraints are available but may increase the number of computationally expensive function evaluations required considerably. As an *ad hoc* alternative, the Hessian matrix was modified whenever estimates out of bounds occurred, making it more diagonally dominant by multiplying its diagonals with a factor, increasing successively until the resulting estimates of covariance matrices were (semi-) positive definite. This made the resulting estimation step more akin to a Method of Steepest Descent step; see Searle *et al.* (1992) for further discussion.

Calculation of  $\log L$  and its derivatives was implemented using a compressed sparse matrix storage scheme as described by George and Liu (1981) and their FORTRAN routines GENQMD and SMBFCT to obtain a minimum degree ordering of  $\mathbf{M}$  and its symbolic factorization. For  $p$  parameters, this required  $1 + p(p + 3)/2$  times as much space as evaluation of  $\log L$  only.

## EXAMPLES

Characteristics of three data sets and the computational requirements for four analyses using both a NR and a DF algorithm (with the same starting values) are summarized in Table 1. The NR algorithm was considered to have converged when the norm of the vector of first derivatives of  $\log L$  was less than  $10^{-4}$  or when the likelihood remained constant to the sixth decimal. The DF analyses used Nelder and Mead's (1965) Simplex procedure to locate the maximum of the likelihood, with a step size of 20% of the starting values given to set up the initial simplex and imposing a value of  $10^{-8}$  for the variance of function values in the simplex as convergence criterion.

For univariate analyses, starting values given were genetic parameters, i.e. heritabilities etc., and one DF evaluation of  $\log L$  was carried out for the NR algorithm to obtain an initial estimate of the error variance, deriving starting values for the other components from it. Preliminary studies had shown the NR algorithm to be sensitive against poor starting values and this strategy to improve convergence, in particular to reduce the incidence of non-permissible estimates in the first iterate markedly. For both models of analysis and starting values close to ('good') and very different to ('bad') the eventual estimates, the NR converged in 5 or 6 iterates. Evaluating 4 first and 10 second partial derivatives of  $\log L$  in the second analysis required about 26 times as long as evaluating  $\log L$  only. Smith (1993) estimated that each first derivative of  $\mathbf{L}$  required twice and each second derivative required four times the work to calculate  $\mathbf{L}$  only, but that using sparse matrix techniques no more than 2–4 times the work would be required for both. However, with only few parameters to be estimated, the NR required considerably longer than the DF algorithm in all four univariate examples shown, even for a relatively detailed model fitting both genetic and permanent environmental effects which are expected to have high negative sampling correlations.

Multivariate analyses were carried out considering all parameters and maximizing with respect to the covariances only initially. For this, variances were fixed to their univariate estimates, for 2 iterates for the NR algorithm and until the variance of function values was less than  $10^{-4}$  for the DF algorithm. For DF analyses, this yielded considerable savings in the number of  $\log L$  to be evaluated, in particular for a trivariate analysis. Again, on average about twice the time required to determine  $\log L$  was needed for each derivative evaluated. With a high dimension of search (9 and 12, respectively, for the bi- and trivariate analysis), the number of function evaluations required for the DF algorithm was large, so that the NR algorithm proved advantageous, requiring less than half of the time needed by the former.

## CONCLUSIONS

Evaluation of first and second partial derivatives of the REML likelihood models is feasible for large scale animal model analyses. Timewise, only about twice the amount required for  $\log L$  only is required per derivative, though additional space needed to store all derivatives of  $\mathbf{M}$  or its Cholesky factor may be considerable. For few parameters to be estimated, there appears to be no advantage in using a NR algorithm over a DF algorithm, though the former will yield an estimate of the covariance matrix of estimates as a by-product. For multivariate analyses, use of the NR algorithm may reduce computational requirements dramatically compared to a DF approach.

However, preliminary experience with the NR algorithm has shown it be markedly less robust than the DF, being sensitive to poor starting values. Furthermore, it is not as easy to constrain estimates to the parameter space. Further research is required to establish optimal use of information from derivatives in locating the maximum of the likelihood, especially when searching at the bounds of the parameter space. Only the use of both first and second

**Table 1** : Examples for Newton-Raphson (NR) versus Derivative-Free (DF) REML

Breed		Hereford		Hereford		Wokalup		Zebu Cross	
		Good <sup>a</sup>	Bad <sup>a</sup>	Good	Bad	All <sup>b</sup>	Sub <sup>b</sup>	All	Sub
Traits(s) <sup>c</sup>		WW		WW		CB, HH		WW, YW, FW	
Effects fitted <sup>d</sup>		<b>a</b>		<b>a, m, c</b>		<b>a, c</b>		<b>a</b>	
Parameters estimated <sup>e</sup>		$\sigma_A^2, \sigma_E^2$		$\sigma_A^2, \sigma_M^2, \sigma_C^2, \sigma_E^2$		$\sigma_{A_{ij}}, \sigma_{C_{ij}}, \sigma_{E_{ij}}$		$\sigma_{A_{ij}}, \sigma_{E_{ij}}$	
No. of parameters		2		4		9		12	
No. of records		3088		3088		2664		2988	
No. of animals		3331		3426		2369		1430	
No. of rows in <b>M</b>		3535		8012		6193		4342	
No. elements in <b>M</b> <sup>f</sup>		100,151		308,060		258,764		135,175	
NR	No. iterates	5	6	5	5	5	6	12	11
	time/iterate [secs]	28.1	28.1	285.5	285.5	371.0	371.0	97.4	97.4
	total time [secs]	143	172	1438	1438	1855	2226	1169	1071
DF	No. log <i>L</i> evaluated	15	23	54	86	699	648	2939	1434
	time/log <i>L</i> [secs]	2.9	2.9	10.7	10.7	7.1	7.1	2.0	2.0
	total time [secs]	44	67	578	920	4963	4620	5878	2868

<sup>a</sup>Univariate analyses for 'good' and 'bad' starting values

<sup>b</sup>Multivariate analyses maximizing w.r.t. all parameters and a subset (covariances only) first

<sup>c</sup>WW : Weaning weight, YW : yearling weight, FW : Final weight, CB : Cannon bone length, HH : Hip height

<sup>d</sup>**a** : direct additive genetic, **m** : maternal genetic, **c** : permanent environmental maternal

<sup>e</sup> $\sigma_{A_{ij}}$  : direct additive genetic,  $\sigma_{M_{ij}}$  : maternal genetic,  $\sigma_{C_{ij}}$  : permanent environmental maternal, and  $\sigma_{E_{ij}}$  : residual (co)variances

<sup>f</sup>Lower triangle only

derivatives has been considered here; the methodology presented, however, puts a whole host of other optimization techniques within our reach; in particular, procedures using first derivatives only (and possibly approximating second derivatives) need to be investigated.

### ACKNOWLEDGMENTS

This work was supported by MRC grant UNE35. I am indebted to S.P. Smith for a preprint of his paper.

### REFERENCES

- DUCROCQ, V. 1993. *Livest. Prod. Sci.* **36** : 143–156.  
 GRASER, H.-U., SMITH, S.P. AND TIER, B. 1987. *J. Anim. Sci.* **64** : 1362–1370.  
 GEORGE, A. AND LIU, J. W.-H. 1981. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632.  
 HARVILLE, D.A. 1977. *J. Amer. Stat. Ass.* **72** : 320–338.  
 MEYER, K. 1989. *Genet. Sel. Evol.* **21** : 317–340.  
 MEYER, K. 1991. *Genet. Sel. Evol.* **23** : 67–83.  
 MISZTAL, I. 1990. *J. Dairy Sci.* **73** : 163–172.  
 MISZTAL, I., LAWLOR, T.L., SHORT, T.H. AND VAN RADEN, P.M. 1992. *J. Dairy Sci.* **75** : 544–551.  
 NELDER, J.A. AND MEAD, R. 1965. *Computer J.* **7** : 147–151.  
 SEARLE, S.R., CASELLA, G. AND MCCULLOCH, C.E. 1992. John Wiley & Sons, Inc., New York.  
 SMITH, S.P. 1993. *J. Comp. Graph. Stat.* : (submitted). .