

Estimating variances and covariances for multivariate animal models by restricted maximum likelihood

K Meyer*

*Edinburgh University, Institute of Animal Genetics,
West Mains Road, Edinburgh EH9 3JN, UK*

(Received 11 May 1988; accepted 17 December 1990)

Summary – Restricted maximum likelihood estimates of variance and covariance components can be obtained by direct maximization of the associated likelihood using standard, derivative-free optimization procedures. In general, this requires a multi-dimensional search and numerous evaluations of the (log) likelihood function. Use of this approach for analyses under an animal model has been described for the univariate case. This model includes animals' additive genetic merit as random effect and accounts for all relationships between animals. In addition, other random factors such as common environmental or maternal genetic effects can be fitted. This paper describes the extension to multivariate analyses, allowing for missing records. A numerical example is given and simplifications for specific models are discussed.

variance component / restricted maximum likelihood / animal model / additional random effect / derivative-free approach / multivariate analysis

Résumé – Estimation par le maximum de vraisemblance restreint (REML) des composantes de variance et de covariance pour un modèle animal multivariable. *En se fondant sur le principe du maximum de vraisemblance restreint, on peut obtenir les estimations des composantes de variance et de covariance par la recherche directe du maximum de la vraisemblance correspondante au moyen de méthodes d'optimisation n'utilisant pas le calcul de dérivées. En général, ceci nécessite une approche multidimensionnelle et de nombreux calculs de la fonction de vraisemblance. L'utilisation de cette approche a déjà été décrite dans le cadre d'un modèle animal avec un seul caractère. Le modèle considère les valeurs individuelles des animaux comme des effets aléatoires, et prend en compte toutes les relations de parenté; de plus, d'autres facteurs de variation aléatoires comme des effets de milieu commun ou des effets maternels génétiques peuvent être pris en compte. Cette étude étend la méthode au cas multivariable et admet que des données soient manquantes. Un exemple numérique est présenté, et les simplifications possibles dans le cas de certains modèles sont discutées.*

composantes de la variance / maximum de vraisemblance restreint / modèle animal / effet aléatoire complémentaire / approche sans dérivation / analyse multivariable

* Present address: AGBU, University of New England, Armidale, NSW 2351, Australia

INTRODUCTION

In the statistical analyses of animal breeding data, traits are often considered one at a time. Usually we are interested, however, not only in the mode of inheritance of a particular trait but also in its relationships with other traits and correlated responses when selecting on the trait analyzed. Multivariate analyses are required to obtain estimates of genetic and phenotypic correlations between traits. Moreover, while univariate analyses implicitly assume that all correlations are 0, joint analyses of correlated traits utilize information from all traits to obtain estimates for a specific trait and are thus likely to yield more accurate results. This is of particular relevance when data are not a random sample, *ie* if records for some traits are missing as the result of selection. For animal breeding data, this is often the case since, typically, data originate from selection experiments or are field records from livestock improvement schemes which select animals on the basis of performance. In that situation, univariate analyses are expected to be biased while multivariate analyses may account for selection.

Analysis of (Co) variance (AOV) type methods have been used widely to estimate genetic and phenotypic correlations. These require records for all traits for all individuals. If there are missing records, this implies that part of the information available is ignored. More importantly, if lack of records is the outcome of selection based on some criterion correlated to trait(s) under analysis, estimates are likely to be biased by selection. In contrast, maximum likelihood (ML) estimation procedures utilize all records available and, under certain conditions, account for selection. Recently this has been considered more formally and from a Bayesian point of view (*eg* Im *et al*, 1989). Even if these conditions are only partially fulfilled, ML estimates are often considerably less biased by selection than their AOV counterparts (Meyer and Thompson, 1984).

A modified ML procedure, so-called restricted maximum likelihood (REML), which accounts for the loss in degrees of freedom due to fixed effects in the model of analysis (Patterson and Thompson, 1971), has become the preferred method of analysis for animal breeding data, not least for its property of reducing selection bias. Multivariate REML algorithms suggested so far, however, in general require the direct inverse of a matrix of size equal to the total number of levels of random effects multiplied by the number of traits considered simultaneously, in each round of an iterative solution scheme. This represents not only a substantial computational requirement but imposes severe limitations on the model and dimension of analysis. Simplifications have only been suggested for the 'equal design matrix' case, *ie* all traits recorded for all animals at the same (or at strictly corresponding) time(s), for models containing only one random factor (*eg* sires) apart from residual errors. To date, there are no practical applications of multivariate REML analyses for models including additional random factors.

REML algorithms as employed in practice today, by and large rely on the use of information from first or even second derivatives of the likelihood function to locate its maximum. Recently, use of a derivative-free procedure, involving explicit evaluation of the likelihood and maximization by direct search has been advocated by Graser *et al* (1987). This was described for univariate analyses fitting animal's additive genetic merit as the only random effect, *ie* estimating 2 variance

components, and required only a 1-dimensional search. Such models, which also incorporate all information on relationships between animals are usually referred to as animal models (AM).

The derivative-free approach provides a flexible and powerful alternative to REML algorithms used currently. Its application for AMs including additional random effects, for instance animals' maternal genetic effects or common environmental effects, for the univariate case has been described previously (Meyer, 1989). This paper presents an extension to multivariate analyses.

THE REML ALGORITHM

The model

Let

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

denote the multivariate linear model of analysis for q traits with: \mathbf{y} the vector of N observations for all traits; \mathbf{b} the vector of NF fixed effects (including any linear or higher order covariables); \mathbf{X} the $N \times NF$ incidence or design matrix for fixed effects with column rank NF^* ; \mathbf{u} the vector of all NR random effects fitted; \mathbf{Z} the $N \times NR$ incidence matrix for random effects; and \mathbf{e} the vector of N random residual errors. Assume that

$$\begin{aligned} V(\mathbf{u}) &= \mathbf{G}, \\ V(\mathbf{e}) &= \mathbf{R} \quad \text{and} \\ \text{Cov}(\mathbf{u}, \mathbf{e}') &= \mathbf{0} \end{aligned}$$

which gives

$$V(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}.$$

Define \mathbf{E} with elements e_{ij} ($i \leq j = 1, \dots, q$) as the symmetric matrix of residual or error covariances between traits. Correspondingly, let $\mathbf{T} = \{t_{ij}\}$ of size $rq \times rq$ denote the matrix of covariances between random effects where r denotes the number of random factors in the model (apart from residual errors). Assume there are $r^* \leq r(r-1)/2$ covariances between the r random factors. The total number of parameters to be estimated is then $s = q(q+1)(r+1)/2 + q^2r^*$.

The likelihood

As outlined previously (Meyer, 1989), the natural log of the likelihood function to be maximised is

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

assuming \mathbf{y} has a multivariate normal distribution with mean $\mathbf{X}\mathbf{b}$ and variance \mathbf{V} , where \mathbf{X}^* (of order $N \times NF^*$) is a full column rank submatrix of \mathbf{X} . Alternatively,

$$-2\log \mathcal{L} = \text{const} + \log |\mathbf{R}| + \log |\mathbf{G}| + \log |\mathbf{C}| + \mathbf{y}'\mathbf{P}\mathbf{y} \quad (3)$$

where \mathbf{C} is the coefficient matrix in the mixed model equations (MME) pertaining to (1) (or a full rank submatrix thereof), and \mathbf{P} is a matrix,

$$\begin{aligned} \mathbf{P} &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \\ &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}^*(\mathbf{X}^{*\prime}\mathbf{V}^{-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{-1} \end{aligned}$$

As described by Graser *et al* (1987), the last 2 terms required in (3), $\log|\mathbf{C}|$ and $\mathbf{y}'\mathbf{P}\mathbf{y}$, can be evaluated simultaneously in a general way for all models of form (1). This involves application of Gaussian Elimination with diagonal pivoting to the matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \quad (4)$$

which is the coefficient matrix in the MME, augmented by the right hand sides and a quadratic in the data vector (see Meyer (1989) for further details).

Calculation of $\log |\mathbf{R}|$

Let \mathbf{y} be ordered according to traits within individuals, and assume error covariances between traits measured on different animals are zero. This results in \mathbf{R} being block-diagonal for animals,

$$\mathbf{R} = \sum_{i=1}^{ND} {}^+ \mathbf{R}_i \quad (5)$$

with ND the number of animals which have records, and \sum^+ denoting the direct matrix sum (Searle, 1982).

For q traits, there are a total of $W = 2^q - 1$ possible combinations of traits recorded. For $q = 2$, for example, $W = 3$ with combinations trait 1 only, trait 2 only and both traits. For animal i which has combination of traits w , \mathbf{R}_i is equal to \mathbf{E}_w , the submatrix of \mathbf{E} obtained by deleting rows and columns pertaining to missing records. This gives

$$\log |\mathbf{R}| = \sum_{w=1}^W N_w \log |\mathbf{E}_w| \quad (6)$$

where N_w represents the number of animals having records for combination of traits w . Hence, evaluation of $\log |\mathbf{R}|$ requires calculation of W log determinants of matrices \mathbf{E}_w of size $q \times q$ or smaller.

Calculation of $\log|\mathbf{G}|$

As for the univariate case, $\log|\mathbf{G}|$ depends on the random effects fitted and their covariance structure. Corresponding conditions to those discussed by Meyer (1989) apply to allow $\log|\mathbf{G}|$ to be evaluated 'indirectly', *ie* without the need to set up and perform numerous Gaussian Elimination steps for a large matrix, to obtain its log determinant, similar to the procedure required to determine $\log|\mathbf{C}|$.

Consider the simplest case with animal's additive genetic merit, denoted by the vector \mathbf{a} of length $q \times NA$ (with NA the total number of animals), as the only random effects in the model, *ie* $r = 1$ and $s = q(q + 1)$. Assume effects are ordered according to traits within animals, and let \mathbf{a}_i be the subvector for the i th animal with covariance matrix \mathbf{T} . Here, \mathbf{T} has dimension $q \times q$ and its elements are the additive genetic covariances. Then,

$$\begin{aligned} \mathbf{u}' &= \mathbf{a}' = (\mathbf{a}'_1 \dots \mathbf{a}'_{NA}) \\ \mathbf{G} &= \mathbf{A} \times \mathbf{T} \end{aligned} \quad (7)$$

$$\log|\mathbf{G}| = NA \log|\mathbf{T}| + q \log|\mathbf{A}| \quad (8)$$

where \mathbf{A} is the numerator relationship matrix between animals, and \times denotes the direct matrix product (Searle, 1982). As for the univariate case, $\log|\mathbf{A}|$ does not depend on the parameters to be estimated and is not required in order to maximize $\log \mathcal{L}$.

Extend the model by allowing for a second random effect for each animal, \mathbf{m} with subvectors $\mathbf{m}_i (i = 1, \dots, NA)$, which has the same correlation structure between animals as \mathbf{a} . A typical example is a maternal genetic effect. Then $r = 2$ and \mathbf{T} is of size $2q \times 2q$, with $s = 3q(q + 1)/2$ if \mathbf{a} and \mathbf{m} are assumed uncorrelated ($r^* = 0$) and $s = q(5q + 3)/2$ otherwise ($r^* = 1$). Assuming \mathbf{u} is ordered according to effects within animals, *ie* $\mathbf{u}' = (\mathbf{a}'_1 \mathbf{m}'_1 \dots \mathbf{a}'_{NA} \mathbf{m}_{NA})$, (7) and (8) hold, with $\mathbf{T} = \text{Var}\{(\mathbf{a}'_i \mathbf{m}'_i)\}$.

Often we want to include an additional random effect, uncorrelated to the other random factors, in the model of analysis. This could be a common environmental effect, such as a litter effect in the analysis of pig data, or the permanent effect due to an animal which is not additive genetic in a multivariate repeatability model. Let this effect, with NC levels per trait, be denoted by \mathbf{c} . \mathbf{T} can then be partitioned into diagonal blocks \mathbf{T}_A (for additive genetic effects) and \mathbf{T}_C (for the additional uncorrelated effect). Correspondingly,

$$\begin{aligned} \mathbf{u}' &= (\mathbf{a}' \mathbf{c}') \\ \mathbf{G} &= \text{Diag} \{ \mathbf{A} \times \mathbf{T}_A; \mathbf{D} \times \mathbf{T}_C \} \end{aligned} \quad (9)$$

$$\log|\mathbf{G}| = NA \log|\mathbf{T}_A| + NC \log|\mathbf{T}_C| + q [\log|\mathbf{A}| + \log|\mathbf{D}|] \quad (10)$$

where \mathbf{D} , most commonly taken to be the identity matrix, describes the correlation structure amongst levels of \mathbf{c} . Again, $\log|\mathbf{A}|$ and $\log|\mathbf{D}|$ are constants and do not need to be evaluated. As above, (9) and (10) also apply, with an appropriate modification to \mathbf{T}_A , if a second random effect, \mathbf{m} , is fitted for each animal. Extensions to other models, for instance including several additional factors \mathbf{c} , can be derived accordingly.

Maximizing the likelihood

Different strategies to locate the maximum of the log likelihood function, or equivalently the minimum of $-2\log\mathcal{L}$, with respect to several parameters have been examined by Meyer (1989). The so-called Simplex procedure of Nelder and Mead (1965) proved to be robust and easy to use and was chosen for the current application, together with the associated convergence criterion of the variance of function values in the Simplex. As for univariate analyses, a step size of 20% was used throughout in setting up the initial Simplex. In particular, this procedure allow constraints on the parameter space to be imposed simply by assigning a very large value to $-2\log\mathcal{L}$ for parameter vectors out of bounds. This is especially important for multivariate analyses, as estimated genetic covariance matrices have a high probability of being non-positive definite, increasingly so with the number of traits considered and the magnitude (absolute value) of genetic correlations (Hill and Thompson (1978)).

To illustrate the convergence behaviour of the maximization procedure, data were simulated, sampling from a multivariate normal distribution, consisting of records for 2 traits for each of 4000 animals, assumed to be offspring of 500 base animals, 100 sires mated to 4 dams each. Fitting an overall mean as the only effect and families ($NC = 400$) as an additional random effect, this gave \mathbf{M} of size 9803 with 80903 non-zero off-diagonal elements. Using the population values for additive genetic variances ($\sigma_{Aij} = 50, 32, 80$ for $i \leq j = 1, \dots, q$), variances due to family or litter effects ($\sigma_{Cij} = 12, 10, 60$) and error variances ($\sigma_{Eij} = 40, 100, 260$) as starting values, 1321 Simplex iterates involving a total of 2437 likelihood evaluations were carried out, at the end of which the variance of function values in the Simplex ($-2\log\mathcal{L}$) was reduced to 1.06×10^{-7} . The behaviour of the multi-dimensional search is illustrated in figures 1 and 2, showing changes in estimates of variance components and associated log likelihoods for successive Simplex iterates. For 'good' starting values, increases in log likelihood after 190 iterates, equivalent to 300 function evaluations and $V(-2\log\mathcal{L}) = 1.9 \times 10^{-4}$, were only very small. Estimates of variances, though, changed until changes in $\log\mathcal{L}$ were of order 10^{-3} or less (see inset of figure 2). About 1000 likelihood evaluations were required to reach that stage of convergence.

SPECIAL CASES

Traits measured on different animals

Specialized multivariate REML algorithms using information from derivatives of the likelihood function have been suggested for models with one random factor for various special cases. Schaeffer *et al* (1978) considered the situation where traits were measured on different animals, so that residual covariances were zero. The algorithm presented here is adapted for this case simply by reducing the vector of parameters to be estimated accordingly. For the analysis under an animal model, however, it has to be borne in mind that with a record for one trait only for each animal, information on genetic covariances is available only through relatives with

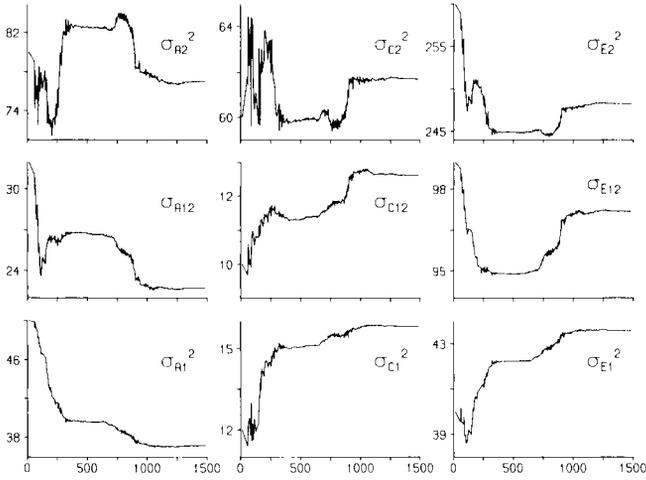


Fig 1. Estimates of variance components (σ_{Aij} additive genetic, σ_{Cij} : common environmental, and σ_{Eij} : residual error, for $i \leq j = 1, 2$) for successive Simplex iterates, plotted against the number of likelihood evaluations performed.

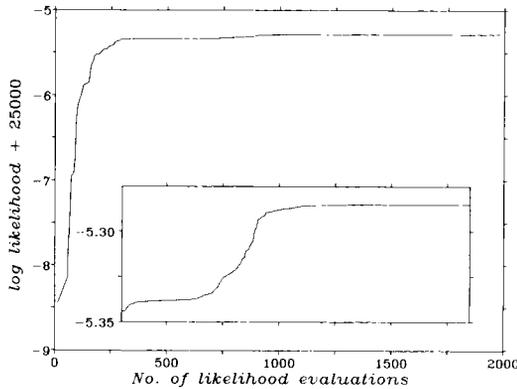


Fig 2. Log likelihoods for successive Simplex iterates, plotted against the number of likelihood evaluations carried out; the inset shows an enlargement of the section for evaluations 300 to 1 750.

records for the other trait(s). Hence estimates are likely to be subject to large sampling errors unless animals are highly related or data sets are large.

Equal design matrices for all traits

If the design or incidence matrices in the linear model are equal for all traits, (1) can be rewritten as

$$\mathbf{y} = (\mathbf{I}_q \times \mathbf{X}_0) \mathbf{b} + (\mathbf{I}_q \times \mathbf{Z}_0) \mathbf{u} + \mathbf{e} \quad (11)$$

assuming, in contrast to the above, that records are ordered according to animals within traits. Let $n = N/q$, $nr = NR/q$ and $nf = NF/q$ denote the number of records, random effects levels and fixed effects levels per trait, respectively. \mathbf{X}_0 of order $n \times nf$ and \mathbf{Z}_0 of order $n \times nr$ are then the design matrices for fixed and random effects for each trait, while \mathbf{I}_q denotes an identity matrix of order q .

Since all animals have records for all traits,

$$\begin{aligned}\mathbf{R} &= \mathbf{E} \times \mathbf{I}_n \\ \log |\mathbf{R}| &= n \log |\mathbf{E}|\end{aligned}\quad (12)$$

Consider now a decomposition of the residual covariance matrix into

$$\mathbf{E} = \mathbf{Q}\mathbf{Q}'$$

This gives

$$\mathbf{R}^{-1} = (\mathbf{Q}^{-T} \times \mathbf{I}_n)(\mathbf{Q}^{-1} \times \mathbf{I}_n)$$

with $\mathbf{Q}^{-T} = (\mathbf{Q}')^{-1}$. Using that

$$(\mathbf{I}_q \times \mathbf{W}')\mathbf{R}^{-1}(\mathbf{I}_q \times \mathbf{W}) = (\mathbf{Q}^{-T} \times \mathbf{I}_{nw})(\mathbf{I}_q \times \mathbf{W}'\mathbf{W})(\mathbf{Q}^{-1} \times \mathbf{I}_{nw})$$

where \mathbf{W} and nw stand in turn for \mathbf{X}_0 and nf , and \mathbf{Z}_0 and nr , it can be shown that \mathbf{R}^{-1} can be factored from the coefficient matrix in the MME.

Transforming the data vector to

$$\mathbf{y}^* = (\mathbf{Q}^{-1} \times \mathbf{I}_n)\mathbf{y}\quad (13)$$

the augmented MME, (4), can be replaced by

$$\mathbf{M}^* = \begin{bmatrix} \mathbf{y}^{*'}\mathbf{y}^* & \mathbf{y}^{*'}(\mathbf{I}_q \times \mathbf{X}_0) & \mathbf{y}^{*'}(\mathbf{I}_q \times \mathbf{Z}_0) \\ (\mathbf{I}_q \times \mathbf{X}_0')\mathbf{y}^* & \mathbf{I}_q \times \mathbf{X}_0'\mathbf{X}_0 & \mathbf{I}_q \times \mathbf{X}_0'\mathbf{Z}_0 \\ (\mathbf{I}_q \times \mathbf{Z}_0')\mathbf{y}^* & \mathbf{I}_q \times \mathbf{Z}_0'\mathbf{X}_0 & \mathbf{I}_q \times \mathbf{Z}_0'\mathbf{Z}_0 + (\mathbf{G}^*)^{-1} \end{bmatrix}\quad (14)$$

with $\mathbf{G}^* = (\mathbf{Q}^{-1} \times \mathbf{I}_{nr})\mathbf{G}(\mathbf{Q}^{-T} \times \mathbf{I}_{nr})$. Absorbing all rows and columns of \mathbf{M}^* into $\mathbf{y}^{*'}\mathbf{y}^*$ then directly yields the quadratic in the data vector required in (3), *ie*

$$\mathbf{y}^{*'}\mathbf{P}^*\mathbf{y}^* = \mathbf{y}'\mathbf{P}\mathbf{y}\quad (15)$$

The log determinant of the coefficient matrix, $\log|\mathbf{C}|$, calculated when operating on \mathbf{M}^* rather than \mathbf{M} , however, has to be adjusted for the fact that \mathbf{R}^{-1} has been factored out.

$$\log|\mathbf{C}| = \log|\mathbf{C}^*| - (nr + nf^*)\log|\mathbf{E}|\quad (16)$$

with $nf^* = NF^*/q$, $\mathbf{C}^* = (\mathbf{Q}' \times \mathbf{I}_{nf+nr})\mathbf{C}(\mathbf{Q} \times \mathbf{I}_{nf+nr})$ and $\mathbf{P}^* = (\mathbf{Q}^{-T} \times \mathbf{I}_n)\mathbf{P}(\mathbf{Q}^{-1} \times \mathbf{I}_n)$. For ease of presentation, (14) has been written for the vector of random effects assumed to be ordered according to effects within traits. For computational purposes, however, some re-ordering would be advisable in order to minimize 'fill-in' during the absorption steps. Ordering effects within animals and animals according to date of birth, for instance, would result in equations for the youngest animals to be eliminated first.

An obvious choice for \mathbf{Q} would be the Cholesky decomposition of \mathbf{E} . Using (14) rather than (4) then reduces computational requirements in setting up the augmented coefficient matrix. However, a large proportion of the off-diagonal elements thus 'saved' initially arise subsequently as fill-in during the absorption steps due to covariances between traits for random effects levels. Alternative forms of \mathbf{Q} exist though which yield \mathbf{G}^* with more or larger diagonal submatrices, *ie* considerably less off-diagonal elements, and are thus computationally advantageous.

Canonical transformation

The use of a canonical transformation of the data to estimate variance components by REML for multivariate linear models with one random factor and equal design matrices for all traits, has been considered by a number of authors. Estimation procedures have been described for expectation-maximization (EM) type algorithms (*eg* Taylor *et al*, 1985; Smith and Graser, 1986) and Fisher's method of scoring (Meyer, 1985).

For q correlated traits, this transformation yields a set of q new traits, so-called canonical variables, which are both genetically and phenotypically uncorrelated. Hence a multivariate analysis can be carried out as a series of q corresponding univariate analyses which results in a substantial reduction of computational effort; see references given for further details.

Consider an AM without additional random effects, *ie* $\mathbf{u} = \mathbf{a}$, $\mathbf{T} = \text{Var}(\mathbf{a}_i)$ and $\mathbf{G} = \mathbf{T} \times \mathbf{A}$, with equal design matrices for all traits. Let λ_i for $i = 1, \dots, q$ denote the eigenvalues of $\mathbf{E}^{-1}\mathbf{T}$ and \mathbf{S} the corresponding matrix of eigenvectors. Then

$$\begin{aligned} \mathbf{STS}' &= \text{Diag} \{ \lambda_i \} & \text{and} \\ \mathbf{SES}' &= \mathbf{I}_q \end{aligned} \quad (17)$$

ie $\mathbf{E} = \mathbf{S}^{-1}\mathbf{S}^{-T}$, $\mathbf{Q} = \mathbf{S}^{-1}$ and $\mathbf{G}^* = \text{Diag} \{ \lambda_i \mathbf{A} \}$. Hence \mathbf{S} , with elements s_{ij} , describes the canonical transformation.

For $\mathbf{y}^* = (\mathbf{S} \times \mathbf{I}_n)\mathbf{y}$, then variance matrix of the transformed data vector, $\mathbf{V}^* = \text{Var}(\mathbf{y}^*)$, the coefficient matrix \mathbf{C}^* and projection matrix \mathbf{P}^* (based on (14)) are block-diagonal for traits (Meyer, 1985). Consequently, \mathbf{M}^* can be partitioned into q independent matrices \mathbf{M}_i^* :

$$\mathbf{M}_i^* = \begin{bmatrix} \mathbf{y}_i^{*'} \mathbf{y}_i^* & \mathbf{y}_i^{*'} \mathbf{X}_0 & \mathbf{y}_i^{*'} \mathbf{Z}_0 \\ \mathbf{X}_0' \mathbf{y}_i^* & \mathbf{X}_0' \mathbf{X}_0 & \mathbf{X}_0' \mathbf{Z}_0 \\ \mathbf{Z}_0' \mathbf{y}_i^* & \mathbf{Z}_0' \mathbf{X}_0 & \mathbf{Z}_0' \mathbf{Z}_0 + \lambda_i^{-1} \mathbf{A}^{-1} \end{bmatrix} \quad (18)$$

where \mathbf{y}_i^* is the subvector of \mathbf{y}^* for trait i .

Clearly, each of these submatrices is equivalent to the augmented coefficient matrix in a univariate analysis of a trait with heritability $\lambda_i/(\lambda_i + 1)$. Absorbing rows and columns 2 to $k = nf + nr + 1$ in (18) (skipping rows with zero pivots) into $\mathbf{y}_i^{*'} \mathbf{y}_i^*$ then yields a quadratic $\mathbf{y}_i^{*'} \mathbf{P}_i^* \mathbf{y}_i^*$ and determinant $\log|\mathbf{C}_i^*|$, where \mathbf{P}_i^* and

\mathbf{C}_i^* are the submatrices of \mathbf{P}^* and \mathbf{C}^* , respectively, for the i -th trait. Quantities required in (3) are then obtained by summing over traits

$$\begin{aligned} \mathbf{y}' \mathbf{P} \mathbf{y} &= \sum_{i=1}^q \mathbf{y}_i^{*'} \mathbf{P}_i^* \mathbf{y}_i^* \\ \log |\mathbf{C}| &= \sum_{i=1}^q \log |\mathbf{C}_i^*| - (nr + nf^*) \log |\mathbf{E}| \end{aligned} \quad (19)$$

Alternatively, the log likelihood can be evaluated as the sum of likelihoods for univariate analyses on the canonical scale together with an adjustment for the transformation.

$$-2 \log \mathcal{L} = -2 \sum_{i=1}^q \log \mathcal{L}_i^* + (n - nf^*) \log |\mathbf{E}| \quad (20)$$

with

$$-2 \log \mathcal{L}_i^* = NA \log \lambda_i + \log |\mathbf{C}_i^*| + \mathbf{y}_i^{*'} \mathbf{P}_i^* \mathbf{y}_i^* \quad (21)$$

Noting that $\mathbf{y}_i^* = \sum_{k=1}^q s_{ik} \mathbf{y}_k$, it follows that

$$\mathbf{y}_i^{*'} \mathbf{P}_i^* \mathbf{y}_i^* = \sum_{k=1}^q \sum_{m=1}^q s_{ik} s_{im} \mathbf{y}_k' \mathbf{P}_i^* \mathbf{y}_m \quad (22)$$

ie that an explicit transformation of the data vector is not required. Replace $\mathbf{y}_i^{*'} \mathbf{y}_i^*$ in (18) by the $q \times q$ matrix of sums of squares and crossproducts between traits on the original scale, $\mathbf{Y} = \{\mathbf{y}_k' \mathbf{y}_m\}$, and expand the first row and column correspondingly, *ie* replace $\mathbf{X}'_0 \mathbf{y}_i^*$ by q columns $\mathbf{X}'_0 \mathbf{y}_k$, and $\mathbf{Z}'_0 \mathbf{y}_i^*$ by columns $\mathbf{Z}'_0 \mathbf{y}_k$, for $k = 1, \dots, q$. Absorbing rows and columns $q+1$ to $q+nf+nr$ into the first q rows and columns then yields $q(q+1)/2$ terms $\mathbf{y}_k' \mathbf{P}_i^* \mathbf{y}_m$, and $\mathbf{y}_i^{*'} \mathbf{P}_i^* \mathbf{y}_i^*$ can be calculated according to (22).

For univariate analyses, the error variance can be estimated directly from the residual sums of squares, *ie* the quadratic in the data vector at the end of the Gaussian Elimination steps (Graser *et al*, 1987). Correspondingly, the error variances on the canonical scale can be determined as

$$\hat{\varepsilon}_i = \mathbf{y}_i^{*'} \mathbf{P}_i^* \mathbf{y}_i^* / (n - nf^*)$$

Back-transforming to the original scale then yields the matrix of residual covariances:

$$\hat{\mathbf{E}} = \mathbf{S}^{-1} \text{Diag}\{\hat{\varepsilon}_i\} \mathbf{S}^{-T}$$

At each iteration, these are the conditional REML estimates of \mathbf{E} given the current value(s) of \mathbf{T} .

This can be utilised to reduce the dimension of search for the maximum of the likelihood function. For the univariate case, Graser *et al* (1987) and Harville and

Callanan (1990) parameterised the analysis to the error variance and its ratio(s) with the other component(s), and maximised the ‘concentrated’ log likelihood, $\log \mathcal{L}_C$ with the respect to the latter only. At the maximum of $\log \mathcal{L}_C$, the estimate of the error variance is equal to its (unconditional) REML estimate and estimates of the other components can be obtained from it and the REML estimates of the variance ratios.

Analogously, we can parameterise the multivariate analysis to \mathbf{E} and a function of \mathbf{E} and \mathbf{T} , $\mathbf{F} = f(\mathbf{E}, \mathbf{T})$. Corresponding to the univariate case, an obvious choice is the generalized ratio of \mathbf{T} and \mathbf{E} . $\log \mathcal{L}_C$ is then maximised with respect to the elements of \mathbf{F} and, at the maximum, $\hat{\mathbf{T}}$ is obtained from $\hat{\mathbf{F}}$ and $\hat{\mathbf{E}}$ via the inverse of $f(\mathbf{E}, \mathbf{T})$. This reduces the dimension of search by q to the q^2 elements of \mathbf{F} . As \mathbf{T} is, by definition, symmetric and is required to set up \mathbf{M} , \mathbf{F} should be chosen so that it is ensured that $\hat{\mathbf{T}}$ is symmetric. A possible strategy is to force \mathbf{F} to be symmetric, eg $\mathbf{F} = \mathbf{E}^{-\frac{1}{2}} \mathbf{T} \mathbf{E}^{-\frac{1}{2}}$ or $\mathbf{F} = \mathbf{L}^{-1} \mathbf{T} \mathbf{L}^{-T}$ with $\mathbf{L} \mathbf{L}' = \mathbf{E}$, and to maximise with respect to the $q(q+1)/2$ elements of the upper triangle of \mathbf{F} only thus reducing the search by a further $q(q-1)/2$.

So far, the canonical decomposition has only been considered as a means to reduce the number of off-diagonal elements in \mathbf{M} and thus the computational requirements in each likelihood evaluation. However, it can be regarded as an alternative parameterization in its own right which effectively reduces the dimension of search to the number of traits. Instead of estimating the $q(q+1)$ elements of \mathbf{T} and \mathbf{E} , we can estimate the q eigenvalues λ_i of $\mathbf{E}^{-1} \mathbf{T}$ and q^2 elements s_{ij} of \mathbf{S} . This is a one-to-one transformation so that maximizing $\log \mathcal{L}$ with respect to elements of $\text{Diag} \{ \lambda_i \}$ and \mathbf{S} is equivalent to maximizing with respect to elements of \mathbf{T} and \mathbf{E} .

Using this parameterization, maximization can be carried out as a nested 2-step procedure, analogous to the approach taken by Smith and Graser (1986) to solve a 2-dimensional problem performing 2 1-dimensional search steps in sequence. From (17), (21) and (22) it follows that (20) can be rewritten as

$$\begin{aligned}
 -2 \log \mathcal{L} = & \sum_{i=1}^q NA \log \lambda_i + \sum_{i=1}^q \log |\mathbf{C}_i^*| \\
 & + \sum_{i=1}^q \sum_{k=1}^q \sum_{m=1}^q s_{ik} s_{im} \mathbf{y}'_k \mathbf{P}_i^* \mathbf{y}_m - (n - n_{f^*}) \log |\mathbf{S}' \mathbf{S}| \quad (23)
 \end{aligned}$$

The first 2 terms in (23) depend on the λ_i only. Hence for given values of λ_i , the log likelihood can be evaluated for different matrices \mathbf{S} . This allows an ‘internal’ search to be performed to determine \mathbf{S} which maximizes $\log \mathcal{L}$ for these λ_i . In general, this search involves q^2 parameters. Computationally it is comparatively undemanding though: once quantities $\log |\mathbf{C}_i^*|$ and $\mathbf{y}'_k \mathbf{P}_i^* \mathbf{y}_m$ ($i = 1, \dots, q$) have been determined, each likelihood evaluation for different values of \mathbf{S} merely requires scalar manipulations or matrix operations proportional to the number of traits.

This ‘internal’ search is repeated for each ‘external’ search step, ie each likelihood evaluation required in maximizing $\log \mathcal{L}$ with respect to the λ_i . Since the dimension of the external search is reduced to q , the number of likelihood evaluations, or equivalently the number of computationally demanding Gaussian Eliminations,

required is lowered substantially. Overall this yielded to a considerable decrease in computational resources needed. After REML estimates λ_i have been determined with the pertaining matrix \mathbf{S} , estimates on the original scale are obtained by reversing the canonical decomposition. This parameterization also has been considered by Juga and Thompson (1990) for 2 traits only, where additional knowledge about the structure of \mathbf{S} could be exploited to reduce the dimensionality of the internal search.

NUMERICAL EXAMPLE

Data from a selection experiment in mice by Sharp *et al* (1984) are given in table I. Records are body weight at 6 wk of age and adjusted food intake from wk 4 to 6 (adjusted for weight at 4 wk by within-family, within-sex regression) for 3 generations of mice selected for increased appetite, *ie* adjusted food intake. In addition, pedigree information for parents of the first generation was available and utilized, yielding a total of 339 animals in the analysis, 284 with records and 55 which were parents only. These data were analyzed for a multivariate animal model with animals as the only random effect (model 1), and for an AM including litters as an additional, uncorrelated random effect (model 2). Fixed effects fitted were generations (3 levels), sex (2 levels) and litter size (7 levels) for each trait, where the first levels of the latter two were set to 0 to account to \mathbf{X} not of full column rank.

Analyses for both models were carried out for data of a general structure (strategy I: \mathbf{M} of form (4)), and accounting for the fact that design matrices for both traits were equal (strategy II: \mathbf{M} of form (14)), utilizing the canonical decomposition of the error (\mathbf{E}) and additive genetic covariance (\mathbf{T}_A) matrices to obtain a transformation of the data which diagonalized the submatrix of \mathbf{G}^* pertaining to animals. In addition, for the analysis under model 1, strategy III used the 'concentrated' likelihood approach, estimating residual covariances directly, while strategy IV used the parameterization to characteristics of the canonical transformation with a nested 2-stage search procedure (\mathbf{M} of form (18) for both).

Preliminary univariate analyses yielded estimates of variance components which were utilized as starting values for the multivariate analyses. Values used were $\sigma_{Aij} = 4.7, 4.0, 8.3$ and $\sigma_{Eij} = 2.5, 3.0, 12.9$ for model 1, and $\sigma_{Aij} = 4.9, 1.0, 6.0, \sigma_{Cij} = 1.5, 1.0, 3.0$, and $\sigma_{Eij} = 1, 7, 1.0, 12.6$ for model 2, for $i \leq j = 1, 2$, respectively. Characteristics of the augmented coefficient matrix \mathbf{M} , components of the log likelihood for the starting values, time required per evaluation of the log likelihood, and estimates of variance components are summarized in table II.

While iterates performed were identical for strategies I and II, the canonical transformation of the data reduced computing time required by more than 50%. For strategy III, a parameterization was chosen so that \mathbf{F} was symmetric ($\mathbf{F} = \mathbf{L}^{-1}\mathbf{T}\mathbf{L}^{-T}$ with $\mathbf{L}\mathbf{L}' = \mathbf{E}$ as described above). Maximizing the likelihood with respect to 3 instead of 6 parameters decreased the number of iterates and likelihood evaluations needed considerably.

However, for this parameterization the search procedure appeared to be less successful in determining the maximum of $\log \mathcal{L}$ accurately. Restarting the search

Table I. Data for body weight at 6 wk (in g) and adjusted food intake (in g), for 3 generations of mice selected for increased appetite, together with pedigree information for parents of generation 1.

F^1	S^2	D^2	L^3	<i>Progeny records⁴</i>							
Base Animals											
101	4R	2E									
103	9P	8C									
104	16T	13C									
105	18Q	17A									
106	23S	22A									
107	25P	24A									
108	30S	29D									
109	2S	4C									
110	6S	7A									
111	8S	9D									
112	13R	16B									
113	17S	18C									
114	22S	23E									
115	24T	25C									
116	29R	30C									
Generation 1											
201	110P	101A	4	A:22.5;59.1	B:22.6;52.4	C:22.9;61.1	D:23.0;57.9	P:24.6;69.3	Q:26.4;66.4	R:24.1;61.6	S:24.3;68.3
203	106R	105F	7	B:23.6;60.1	C:24.2;60.8	D:22.5;62.8	F:22.3;59.4	P:23.7;66.9	Q:27.5;72.9	R:17.3;58.4	
204	108Q	107A	2	A:20.3;58.1	B:21.3;58.7	C:21.9;53.2	D:22.1;57.1	Q:28.8;62.5			
205	110P	109B	7	A:18.9;59.3	C:23.2;69.0	D:23.8;64.6	E:27.9;60.7	P:21.6;68.4	Q:25.9;73.0	R:27.6;73.2	
206	112U	111C	7	A:24.4;63.6	C:21.7;62.6	E:19.7;62.1	G:19.9;63.8	P:18.6;70.6	Q:25.4;76.3	R:20.2;76.6	S:23.6;73.5
208	116R	115E	5	A:17.3;61.0	B:17.2;60.7	C:14.6;59.3	D:17.3;59.4	P:19.2;70.9	Q:18.8;73.4	R:18.0;68.9	S:24.0;71.8
210	103Q	104D	4	A:19.2;56.5	C:23.8;65.5	D:22.5;58.2	F:23.1;55.4	P:27.1;65.4	Q:34.5;65.1	R:28.1;67.7	
211	105P	106B	5	A:25.5;60.6	B:24.1;54.9	C:27.3;56.0	P:26.7;60.2	Q:31.5;66.4	S:25.3;60.8	T:29.9;64.8	
212	107S	108A	4	A:23.8;58.3	B:20.1;60.9	C:21.5;52.7	P:24.6;60.2	Q:27.6;68.1	S:21.8;57.5	T:25.2;62.7	
213	109U	110B	5	A:25.8;62.2	B:28.1;60.1	C:25.5;62.6	D:23.1;60.0	P:28.8;66.9	Q:33.8;70.8	R:31.6;68.9	S:30.3;75.0
214	111S	112D	4	A:23.7;62.7	B:22.9;62.4	C:21.8;58.4	P:24.2;61.7	Q:23.6;61.2	T:24.9;65.1	U:19.5;60.9	
215	113T	114D	2	A:22.5;59.2	B:23.4;55.9	C:21.4;57.8	D:22.7;57.5	P:26.3;66.1	Q:24.5;61.3		
216	115S	116C	5	A:21.7;53.6	B:23.8;61.3	C:18.1;52.8	D:22.0;61.0	P:25.2;60.7	R:26.8;63.8	S:26.4;62.2	T:26.0;68.5
Generation 2											
302	204Q	203D	6	A:23.9;59.6	B:24.3;53.3	C:23.5;55.2	D:21.7;54.6	P:24.2;64.7	Q:26.8;67.3	S:26.2;65.6	T:23.1;65.1
303	206R	205C	1	A:24.5;55.1	B:26.0;52.1	C:24.4;54.0	P:26.7;63.6				
304	208Q	215C	5	A:16.0;58.8	B:18.8;64.2	C:17.3;60.0	D:19.7;58.0	P:21.5;70.4	Q:20.0;68.7	R:20.3;69.9	S:19.1;66.6
306	212Q	211A	5	A:25.4;59.0	B:24.4;60.0	C:25.7;58.4	D:26.7;59.7	P:30.3;69.4	Q:31.4;68.2		
307	214T	213C	5	A:21.9;65.7	B:25.2;63.8	C:22.2;62.7	E:22.0;63.0	Q:26.1;71.5	R:26.4;78.0	S:23.7;70.4	T:23.3;69.6
308	216T	215A	3	A:21.5;48.6	B:20.9;61.6	P:25.8;75.4	R:26.1;69.7	S:26.2;71.5	U:28.0;76.4		
309	201S	210D	2	A:23.3;56.9	B:25.2;69.1	D:24.2;62.6	P:23.6;67.6	Q:28.2;66.1	R:25.9;67.5		
311	205R	206G	3	A:19.6;70.9	C:22.4;67.9	E:23.2;71.0	F:22.5;61.8				
312	215Q	208A	2	A:20.0;53.4	B:18.7;46.9	C:19.5;59.1	P:21.1;60.8	Q:24.7;69.1	R:22.2;59.0		
313	201P	210C	5	A:23.1;57.0	B:22.9;63.8	D:22.1;60.6	F:25.7;59.9	P:27.1;73.8			
314	211Q	212B	5	A:21.4;64.4	B:19.9;64.3	C:22.9;61.3	D:20.7;63.8	P:24.2;68.0	R:25.0;65.7	S:27.0;58.2	T:27.2;63.4
315	213S	214A	5	A:22.2;63.3	B:18.4;59.3	D:20.6;62.3	E:18.9;62.0	P:23.3;76.3	Q:28.9;76.2	R:24.4;78.1	
316	215P	216B	7	A:18.9;62.4	B:18.8;65.0	C:18.1;61.5	D:18.8;60.0	P:21.4;73.5	Q:19.9;69.8	R:18.6;68.7	T:21.2;73.5
Generation 3											
401	302Q	309D	6	A:24.6;64.6	B:22.6;61.7	C:25.5;65.5	Q:24.7;66.2	S:24.9;66.9	T:25.6;64.2	U:26.0;66.9	
402	304P	303A	6	A:22.7;70.3	B:24.4;70.1	C:23.8;73.5	D:23.9;77.8	Q:29.1;74.0	R:27.0;67.8	T:27.1;73.0	
403	306P	313D	5	A:24.5;66.6	B:25.3;71.4	C:24.4;68.5	D:22.9;65.2	P:26.4;78.0	Q:26.0;77.0	R:24.4;72.5	U:27.5;74.7
404	308U	307A	6	A:19.9;64.0	B:17.5;59.8	C:20.3;66.8	D:22.1;68.9	Q:22.8;69.0	R:25.7;70.2	T:30.5;82.1	U:24.8;73.1
405	302S	309B	6	A:27.2;69.1	B:23.0;80.4	C:26.6;55.3	P:28.0;70.0	Q:26.2;66.8	R:20.8;56.6	S:26.4;58.4	
406	312P	311E	6	A:22.4;65.0	B:22.0;64.2	C:23.0;65.0	E:20.5;66.7	Q:26.8;74.5	R:22.3;74.2	S:26.6;66.4	T:24.0;66.5
407	314P	313B	2	A:21.9;59.8	B:21.3;57.3	C:22.7;62.4	P:27.1;63.5	Q:26.3;66.3	R:24.6;59.6		
408	316P	315A	1	P:31.5;61.6	Q:33.8;58.5						
409	309R	302A	5	B:25.2;69.1	D:24.8;73.7	E:22.8;64.2	F:23.2;63.7	P:30.4;64.6	Q:29.9;65.4	R:31.0;62.2	
410	303P	304B	1	A:26.1;62.8	B:24.0;64.0	C:20.5;51.3	D:27.7;59.3	P:27.5;59.6			
411	309R	306B	2	A:24.3;61.3	B:25.5;59.8	C:22.7;60.0	D:22.1;57.4	P:25.3;68.6	Q:31.1;67.4		
412	307R	308B	3	A:26.0;65.7	B:26.2;65.4	C:26.3;59.6	P:24.3;61.7	R:28.6;64.4	S:25.7;58.2	T:27.8;67.4	
413	309P	302C	4	A:24.1;65.1	B:21.4;59.3	D:24.2;57.1	E:23.1;56.7	P:24.1;57.7	Q:26.0;68.6	R:28.2;67.9	
414	307Q	312C	2	A:24.2;63.6	B:23.1;59.5	P:24.0;69.4	Q:24.5;67.2	R:27.0;65.6	S:23.1;70.5		
415	313P	314A	5	A:23.7;64.3	B:23.7;64.5	C:23.7;68.0	D:26.2;67.1	P:28.2;65.0	Q:26.6;65.5	T:31.0;71.4	U:30.4;67.7
416	315R	316B	3	A:23.6;62.6	B:25.5;67.7	C:23.7;62.1	D:25.1;62.8	P:23.3;66.6	Q:26.0;60.6	R:27.0;65.2	S:26.3;60.7

¹ Family ID, coded as '(generation No + 1) × 100 + family No within generation'. ² Animal ID, coded as 'family ID × 10 + animal ID within family', where within family IDs are A, B, C,... for females, and P, Q, R,... for males. ³ Litter size to which litter was trimmed after birth (1: 5 or less; 2: 6 or 7; 3: 8; 4: 9; 5: 10; 6: 11; 7: 12 pups). ⁴ Given as: 'Animal ID (within family)': '6-wk weight'; 'adjusted food intake', for up to 8 progeny per family (this was the maximum number recorded per family).

from the ‘converged’ values eventually led to the same estimates as for strategies I and II, but this required several restarts and an excessive number of likelihood evaluations, indicating problems of numerical nature, so that, on the whole strategy III appeared less advantageous than II. Strategy IV performed markedly better. The internal search for \mathbf{S} which maximized $\log \mathcal{L}$ for given values of λ_i increased the time per likelihood evaluation by $\approx 5\%$, but reduced the number of Gaussian Elimination steps to a fraction of those required by the other strategies and reduced the total time for the analysis to $< 3\%$ of that required when ignoring the special structure of the data (strategy I).

CONCLUSIONS

The derivative-free approach is well suited to the multivariate estimation of variance components by REML. It provides a flexible algorithm which can be adapted to a range of models, differing in the random effects fitted and assumptions about covariances between them, of interest for the analysis of animal breeding data. As for multivariate REML algorithms described previously though, computational requirements are high. Not only is the time required for each likelihood evaluation increased considerably in comparison to the univariate case, but the number of parameters to be estimated simultaneously rises markedly with the number of traits. This causes a dramatic increase in the number of search steps needed to locate the maximum of the likelihood, with a correspondingly high number of likelihood evaluations required.

Components of the likelihood comprise the determinants of the covariance matrices of the random effects fitted and of the residual errors, as well as the determinant of the coefficient matrix in mixed model equations and a quadratic function of the data vector. The first 2 can often be obtained indirectly, requiring only the determinants of matrices of variance components of size $qr \times qr$, $q \times q$ or less. Calculation of the latter 2 terms, however, requires a matrix of size proportional to the number of traits and the number of levels of fixed and random effects in the model, to be set up and ‘swept out’. This matrix is typically very large but also sparse so that sparse matrix techniques can be employed successfully to handle analyses involving several thousand effects (see Graser *et al* (1987) or Meyer (1989) for details).

Computational demands to absorb all rows and columns of the augmented coefficient matrix into the first element are determined not only by the size of the matrix but, more importantly, by the number of non-zero off-diagonal elements. For multivariate analyses, this number is generally large due to contributions from covariances between traits. If design matrices are equal for all traits, a canonical transformation of the data can be utilized to eliminate a considerable proportion of these off-diagonal elements, thus reducing computational effort required for each likelihood evaluation markedly. Furthermore, if the model of analysis contains only one random factor, we can reparameterize our analysis from the covariance components to the eigenvalues and elements of the corresponding eigenvectors arising in the canonical decomposition. This allows maximization to be carried out as a nested 2-step procedure, reducing computational requirements dramatically.

Table II. Components of the log likelihood and estimates of variances due to animals additive genetic effects (σ_{Aij}), litters (σ_{Cij}) and residual errors (σ_{Eij}), for data given in table I (trait 1: 6 wk weight, trait 2: adjusted food intake) and for models of analysis ignoring litters (model 1) and including litter effects (model 2), for alternative forms of the augmented coefficient matrix (\mathbf{M}) in the mixed model equations (I: general form ignoring special data structure; II: utilize that design matrices are equal for all traits; III: special form for models with animals as the only random effect and equal design matrices using a canonical transformation and estimating error covariances directly, IV: as III but reparameterizing to the components of the canonical decomposition and carrying out maximization as a nested 2-stage procedure); all logarithmic values given are to base e .

Computing strategy	Model 1				Model 2	
	I	II	III	IV	I	II
Dimension of search	6	6	3	2(+4)	9	9
M of form [Equ. No.]	(4)	(14)	(18)	(18)	(4)	(14)
No. of equations in M	703	703	2× 353	2× 353	787	787
No. off-diag. coeff. in M	7195	3718	2× 1481	2× 1481	9121	4744
Time per log \mathcal{L}^1	1.985	0.654	0.682	0.715	7.742	2.887
<i>Components of log \mathcal{L} for starting values</i>						
$\mathbf{y}'\mathbf{R}^{-1}\mathbf{y}^2$	2357.820	2357.820			2572.058	2572.058
$\mathbf{y}'\mathbf{P}\mathbf{y}^{(2)}$	690.140	690.140	314.771	274.005 ⁴	546.262	546.262
			+ 375.370	274.187		
log $ \mathbf{C} ^{(2,3)}$	-295.087	802.973	238.573	238.577	-318.843	860.615
			+ 564.400	564.400		
log $ \mathbf{E} $	3.1463		3.1463	3.4430	3.0165	
log $ \mathbf{T}_A $	3.1359		-0.01038	-0.01038	3.3464	
log $ \mathbf{T}_C $					1.2528	
log $\mathcal{L} + 1100$	-75.841814		-75.841814	-45.51546	-35.575795	
<i>Estimates for $\text{Var}(-2 \log \mathcal{L}) \leq 10^{-8}$</i>						
No. of Simplex iterates	196		79	12	481	
No. of log \mathcal{L} evaluated	325		181	26	722	
$\text{Var}(-2 \log \mathcal{L})[\times 10^{-9}]$	5.24		7.47	6.96	4.99	
Total time (CPU seconds)	645		212	123	17	
log $\mathcal{L} + 1100$	-45.482172		-45.755523	-45.482097	-30.069189	
σ_{A11}	4.376		4.017	4.369	4.990	
σ_{A12}	0.165		-0.252	0.149	-0.476	
σ_{A22}	7.926		8.584	7.910	6.387	
σ_{C11}					1.517	
σ_{C12}					-0.752	
σ_{C22}					3.081	
σ_{E11}	2.618		2.920	2.619	1.633	
σ_{E12}	2.065		2.420	2.064	2.757	
σ_{E22}	13.096		12.941	13.088	12.430	

¹ CPU time in seconds required per likelihood evaluation. ² Components for separate univariate analysis on the canonical scale for strategy III. ³ $\log|\mathbf{C}^*|$ for strategies II and III. ⁴ Values given are for \mathbf{S} which maximizes log \mathcal{L} (after 78 internal iterates); values for starting values of \mathbf{S} are identical to those given under category III.

Further research is required to investigate potential improvements of the algorithm presented. There may be alternative parameterizations which give quicker convergence, *ie* for which the maximum of the likelihood can be determined more easily. The Simplex method has been reported to perform well for few dimensions but to be progressively less successful as the dimension of search increases (Box, 1966). Use of a NAG library subroutine (E04CCF) which incorporates modifications of the Simplex procedure, as suggested by Parkinson and Hutchinson (1972), to improve its efficiency failed to reduce the number of likelihood evaluations required for the cases examined (models 1 and 2, 2 traits, general data structure). Other optimization procedures may require less iterates and thus reduce the number of likelihood evaluations necessary.

In particular, it should be noted that by expanding the quadratic in the data vector to the corresponding matrix of weighted sums of squares and crossproducts for multiple right hand sides, as described in the *Canonical transformation* section, each Gaussian Elimination step can yield several points on the likelihood surface. While the Simplex method used here is sequential, an alternative search strategy might exploit this additional information and thus reduce the heavy computational demands of a multivariate animal model analysis.

ACKNOWLEDGMENTS

This study was supported by the Agricultural and Food Research Council. I am grateful to WG Hill for comments and R Thompson for explaining the concept of Juga and Thompson. A referee emphasized the potential scope of simultaneous multiple likelihood evaluations through multiple right hand sides.

REFERENCES

- Box MJ (1966) A comparison of several current optimization methods, and the use of transformations in constrained problems. *Computer J* 9, 67-77
- Graser HU, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood *J Anim Sci* 64, 1362-1370
- Harville DA, Callanan TP (1990) Computational aspects of likelihood-based inference for variance components. *In: Proc Int Symp Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer Verlag, Heidelberg
- Hill WG, Thompson R (1978) Probabilities of nonpositive definite between group or genetic covariance matrices. *Biometrics* 34, 429-439
- Juga J, Thompson R (1990) Estimation of bivariate variance components. *In: Proc 4th Wld Cong Genet Appl Livest Prod* (Hill WG, Thompson R, Wolliams JA, eds) vol XIII 496-499
- Im S, Fernando RL, Gianola D (1989) Likelihood inferences in animal breeding under selection: a missing data theory view point. *Genet Sel Evol* 21, 399-414
- Meyer K (1985) Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41, 153-165

- 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, 317-340
- Meyer K, Thompson R (1984) Bias in variance and covariance component estimators due to selection on a correlated trait. *J Anim Breed Genet* 101, 33-50
- Numerical Algorithms Group (1988) *Nag Reference Manual, Mark 11*
- Nelder JA, Mead R (1965) A simplex method for function minimization. *Computer J* 7, 147-151
- Parkinson JM, Hutchinson D (1972) An investigation into the efficiency of variants of the Simplex method. In: *Numerical Methods for Non-Linear Optimization*. Academic Press, London, 115-135
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-554
- Schaeffer LR, Wilton JW, Thompson R (1978) Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* 34, 199-208
- Searle SR (1982) *Matrix Algebra Useful for Statistics*. John Wiley & Sons, NY
- Sharp GL, Hill WG, Robertson A (1984) Effects of selection on growth, body composition, and food intake in mice. I. Response in selected traits. *Genet Res* 43, 75-92
- Smith SP, Graser HU (1986) Estimating variance components in a class of mixed models by restricted maximum likelihood. *J Dairy Sci* 69, 1156-1165
- Taylor JF, Bean B, Marshall CE, Sullivan JJ (1985) Genetic and environmental components of semen production traits of artificial insemination Holstein bulls. *J Dairy Sci* 68, 2703-2722