An "average information" Restricted Maximum Likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices

K. Meyer

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

Summary

A Quasi-Newton Restricted Maximum Likelihood algorithm which approximates the Hessian matrix with the average of observed and expected information is described for the estimation of covariance components or covariance functions under a linear mixed model. The computing strategy outlined relies on sparse matrix tools and automatic differentiation of a matrix, and does not require inversion of large, sparse matrices.

For the special case of a model with only one random factor and equal design matrices for all traits, calculations to evaluate the likelihood, first and 'average' second derivatives can be carried out trait by trait, collapsing computational requirements of a multivariate analysis to those of a series of univariate analyses. This is facilitated by a canonical decomposition of the covariance matrices and corresponding transformation of the data to new, uncorrelated traits.

The rank of the estimated genetic covariance is determined by the number of non-zero eigenvalues of the canonical decomposition, and thus can be reduced by fixing a number of eigenvalues at zero. This limits the number of univariate analyses needed to the required rank. It is particularly useful for the estimation of covariance function when a potentially large number of highly correlated traits can be described by a low order polynomial.

Keywords : REML, average information, covariance components, reduced rank, covariance function, equal design matrices

INTRODUCTION

Estimation of (co)variance components by Restricted Maximum Likelihood (REML) fitting an animal model to date is mainly carried out using a derivative-free (DF) algorithm as initially proposed by Graser *et al.* (1987). While this has been found to be slow to converge, especially for multi-trait and multi-parameter analyses, it does not require the inverse of a large matrix and can be implemented efficiently using sparse matrix storage and factorisation techniques, making it computationally feasible for models involving tens of thousands of animals.

Recently there has been renewed interest in algorithms utilising derivatives of the likelihood function to locate its maximum. This has been furthered by technical advances, making computations faster and allowing larger and larger matrices to be stored. Moreover, the rediscovery of Takahashi *et al.*'s (1973) algorithm to invert large sparse matrices has removed most of the constraints on algorithms imposed previously by the need to invert large matrices.

In particular, "average information" (AI) REML, a Quasi-Newton algorithm which requires first derivatives of the likelihood but replaces second derivatives with the average of the observed and expected information, described by Johnson and Thompson (1995) has been found to be computationally highly advantageous over DF procedures.

It is well recognised that for several correlated traits, most information available is contained in a subset of the traits or linear combinations thereof. This subset is the smaller the higher the correlations between traits. More technically, several eigenvalues of the corresponding covariance matrix between traits are very small or zero. If

1

a modified covariance matrix were obtained by setting all small eigenvalues to zero and backtransforming to the original scale (using the eigenvectors corresponding to non-zero eigenvalues), it would have reduced rank.

There has been interest in reduced rank covariance matrices in several areas. Wiggans *et al.* (1995; unpublished) collapsed the multivariate genetic evaluation for 30 traits (10 test day records each for milk, fat and protein yield in dairy cows) to the equivalent of 5 univariate analyses by reducing the rank of the genetic covariance matrix and exploiting a transformation to canonical scale. Kirkpatrick and Heckman (1989) introduced the concept of 'covariance functions', expressing the covariance between traits as a higher order polynomial function. Polynomials can be fitted to full or reduced order. In the latter case, the resulting covariance matrix has reduced rank, i.e. a number of zero eigenvalues (Kirkpatrick *et al.*, 1990).

The covariance function (CF) model was developed with the analysis of "traits" with potentially infinitely many repeated, or almost repeated records in mind, where the phenotype or genotype of individuals is described by a function rather than a finite number of measurements (Kirkpatrick and Heckman, 1989). A typical example is growth curve of an animal. Hence, in essence, CFs are the infinite-dimensional equivalent of covariance matrices. Analysis under a CF model implies that coefficients of the CF are estimated rather than individual covariances as under the usual multivariate, "finite" linear model; see Kirkpatrick *et al.* (1990) for further details.

While it is possible to modify an estimated covariance matrix to reduce its rank (as done by Kirkpatrick *et al.* (1990 and 1994)), it would be preferable to impose restrictions on the rank of covariance matrices 'directly' during (REML) estimation. Ideally, this could be done increasing the order of fit (i.e. rank allowed) sequentially until an additional non-zero eigenvalue does not increase the likelihood significantly.

Conceptually, this could be implemented simply by reparameterising, to the eigenvalues and corresponding eigenvectors of a covariance matrix, and fixing the required number of eigenvalues at zero. Practical applications of such reparameterisations, however, have been restricted to simple animal models with equal design matrices for all traits; see Jensen and Mao (1988) for a review. For these, a canonical decomposition of the genetic and residual covariance matrix together yields a transformation to uncorrelated variables with unit residual variance, leaving the number of parameters to be estimated unchanged (for full rank).

Meyer and Hill (1997) described how REML estimates of CFs or, more precisely, their coefficients could be obtained using a DF algorithm through a simple reparameterisation of the variance component model. However, they found it slow to converge for orders of fit greater than 3 or 4. Moreover, for simulated data sets the DF algorithm failed to locate the maximum of the likelihood accurately in several instances, especially if CFs were fitted to a higher order than simulated.

This paper reviews an AI-REML algorithm for the general, multivariate case, presenting a computing strategy which does not require sparse matrix inversion. Subsequently, simplifications for the special case of a simple animal model with equal design matrices for all traits are considered. Additional reductions in computational requirements are shown for the estimation of reduced rank genetic covariance matrices or reduced order CFs.

THE GENERAL CASE

Model of analysis

Consider the multivariate linear mixed model for t traits

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1}$$

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

For an animal model, $\mathbf{u}$ always includes the vector of animals' additive genetic effects ($\mathbf{a}$). In addition, it may contain other random effects, such as animals' maternal genetic effects, permanent environmental effects due to the animal or its dam, or common environmental effects such as litter effects.

Let $\Sigma_A = \{\sigma_{A_{ij}}\}$, denote the $t \times t$ matrix of additive genetic covariances. For $\mathbf{u} = \mathbf{a}$ this gives $\mathbf{G} = \Sigma_A \otimes \mathbf{A}$ where $\mathbf{A}$ is the numerator relationship matrix and $'\otimes'$ denotes the direct matrix product. If other random effects are fitted, $\mathbf{G}$

is expanded correspondingly; see Meyer (1991) for a more detailed description. Assuming $\mathbf{y}$ is ordered according to traits within animals,

$$\mathbf{R} = \sum_{i=1}^{N} {}^{+}\mathbf{R}_i \tag{2}$$

where is N the number of animals which have records, and $\sum^{+}$ denotes the direct matrix sum (Searle, 1982). Let $\Sigma_E = \{\sigma_{E_{ij}}\}$ be the matrix of residual covariances between traits. For t traits, there are a total of $W = 2^t - 1$ possible combinations of traits recorded (assuming single records per trait), e.g. $W = 3$ for $t = 2$. For animal i with combination of traits w, $\mathbf{R}_i$ is equal to Σ_{Ew}, the submatrix of Σ_E obtained by deleting rows and columns pertaining to missing records.

Average information REML

Assuming a multivariate normal distribution, i.e. $\mathbf{y} \sim N(\mathbf{Xb}, \mathbf{V})$, the log of the REML likelihood ($\mathcal{L}$) is (e.g. Harville 1977)

$$\log \mathcal{L} = -\frac{1}{2} \left[const + \log |\mathbf{V}| + \log |\mathbf{X}^{*'}\mathbf{V}^{-1}\mathbf{X}^*| + \mathbf{y}'\mathbf{Py} \right] \tag{3}$$

where $\mathbf{X}^*$ denotes a full-rank submatrix of $\mathbf{X}$, and $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}^*(\mathbf{X}^{*'}\mathbf{V}^{-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{V}^{-1}$.

Let θ denote the vector of parameters to be estimated with elements θ_i for $i = 1, \ldots, p$. Derivatives of $\log \mathcal{L}$ are then (Harville, 1977)

$$\frac{\partial \log \mathcal{L}}{\partial \theta_i} = tr\left(\frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{P} \right) - \mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{Py} \tag{4}$$

and

$$\frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} = -\frac{1}{2} \left[tr(\mathbf{P}\frac{\partial^2 \mathbf{V}}{\partial \theta_i \partial \theta_j}) - tr(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_j}) - \mathbf{y}'\mathbf{P}(\frac{\partial^2 \mathbf{V}}{\partial \theta_i \partial \theta_j} - 2\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_j})\mathbf{Py} \right] \tag{5}$$

The latter is commonly called the *observed information*. It has expectation

$$E\left[\frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} \right] = -\frac{1}{2}tr\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_j} \right) \tag{6}$$

For $\mathbf{V}$ linear in θ, $\partial^2 \mathbf{V}/\partial \theta_i \partial \theta_j = 0$, and the *average* of observed (5) and expected (6) *information* is (Johnson and Thompson, 1995)

$$\frac{1}{2}\left(\frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} + E\left[\frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} \right] \right) = -\frac{1}{2}\mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{Py} \tag{7}$$

The right hand side of (7) is (except for a scale factor) equal to the second derivative of $\mathbf{y}'\mathbf{Py}$ with respect to θ_i and θ_j, i.e. the average information is equal to the data part of the observed information.

REML estimates of θ can then be obtained by substituting the average information matrix for the Hessian matrix in a suitable optimisation scheme which uses information from second derivatives of the function to be maximised; see Meyer and Smith (1996) for a detailed discussion of Newton-Raphson type algorithms in this context.

Calculation of the log likelihood

Calculation of $\log \mathcal{L}$ pertaining to (1) has been described in detail by Meyer (1991). It relies on rewriting (3) as (Graser *et al*. 1987; Meyer 1989)

$$\log \mathcal{L} = -\frac{1}{2} \left[const + \log |\mathbf{R}| + \log |\mathbf{G}| + \log |\mathbf{C}| + \mathbf{y}'\mathbf{Py} \right] \tag{8}$$

3

where $\mathbf{C}$ is the coefficient matrix in the mixed model equations (MME) for (1) (or a full rank submatrix thereof).

The first two components of $\log \mathcal{L}$ can usually be evaluated indirectly, requiring only the log determinants of matrices of size equal to the maximum number of records or effects fitted per animal. For $\mathbf{u} = \mathbf{a}$,

$$\log |\mathbf{G}| = N_A \log |\Sigma_A| + t \log |\mathbf{A}| \tag{9}$$

where N_A denotes the number of animals in the analysis (including parents without records). $\log |\mathbf{A}|$ is a constant and can be omitted for the purpose of maximising $\log \mathcal{L}$. Similarly with N_w denoting the number of animals having records for combination of traits w

$$\log |\mathbf{R}| = \sum_{w=1}^{W} N_w \log |\Sigma_{Ew}| \tag{10}$$

The other two terms in (8), $\log |\mathbf{C}|$ and $\mathbf{y}'\mathbf{Py}$, can be determined in a general way for all models of form (1). Let $\mathbf{M}$ (of size $M \times M$) denote the mixed model matrix (MMM), i.e. the coefficient matrix in the MME augmented by the vector of right hand sides ($\mathbf{r}$) and a quadratic in the data vector.

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{r} \\ \mathbf{r}' & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{11}$$

A Cholesky decomposition of $\mathbf{M}$ gives $\mathbf{M} = \mathbf{L}\mathbf{L}'$, with $\mathbf{L}$ a lower triangular matrix with elements l_{ij} ($l_{ij} = 0$ for $j > i$), and

$$\log |\mathbf{C}| = 2 \sum_{k=1}^{M-1} \log l_{kk} \tag{12}$$

$$\mathbf{y}'\mathbf{Py} = l_{MM}^2 \tag{13}$$

Factorisation of $\mathbf{M}$ for large scale animal model analyses is computationally feasible through the use of sparse matrix techniques; see, for instance, George and Liu (1981).

Calculation of first derivatives

Differentiating (8) gives partial first derivatives

$$\frac{\partial \log \mathcal{L}}{\partial \theta_i} = -\frac{1}{2} \left[\frac{\partial \log |\mathbf{R}|}{\partial \theta_i} + \frac{\partial \log |\mathbf{G}|}{\partial \theta_i} + \frac{\partial \log |\mathbf{C}|}{\partial \theta_i} + \frac{\partial \mathbf{y}'\mathbf{Py}}{\partial \theta_i} \right] \tag{14}$$

Analogously to the calculation of $\log \mathcal{L}$ the first two terms in (14) can usually be determined indirectly while the other two terms can be evaluated extending the Cholesky factorisation of the MMM (Meyer and Smith, 1996).

Let $\mathbf{D}_A^{\theta_i} = \partial \Sigma_A / \partial \theta_i$ be a matrix whose elements are 1 if θ_i is equal to the $kl-$th element of Σ_A and zero otherwise. Further, let δ_{kl} denote Kronecker's Delta, i.e., $\delta_{kl} = 1$ for $k = l$ and zero otherwise, and σ_A^{kl} denotes the kl-th element of Σ_A^{-1}. For $\theta_i = \sigma_{A_{kl}}$

$$\frac{\partial \log |\mathbf{G}|}{\partial \sigma_{A_{kl}}} = N_A \, tr(\Sigma_A^{-1} \mathbf{D}_A^{\sigma_{A_{kl}}}) = N_A (2 - \delta_{kl}) \sigma_A^{kl} \tag{15}$$

Similarly, with $\mathbf{D}_w^{\theta_i} = \partial \Sigma_{Ew} / \partial \theta_i$ and σ_{Ew}^{kl} the kl-th element of Σ_{Ew}^{-1},

$$\frac{\partial \log |\mathbf{R}|}{\partial \sigma_{E_{kl}}} = \sum_{w=1}^{W} N_w tr(\Sigma_{Ew}^{-1} \mathbf{D}_w^{\sigma_{E_{kl}}}) = \sum_{w=1}^{W} N_w (2 - \delta_{kl}) \sigma_{Ew}^{kl} \tag{16}$$

while all other first derivatives of $\log |\mathbf{G}|$ and $\log |\mathbf{R}|$ are zero.

Smith (1995) describes a procedure for automatic differentiation of the Cholesky decomposition. In essence, it is an extension of the Cholesky factorisation which gives not only the Cholesky factor of a matrix but also its derivatives, provided the corresponding derivatives of the original matrix can be specified. In particular, Smith (1995) outlines a 'backwards differentiation' scheme which is applicable when we want to evaluate a scalar function of $\mathbf{L}$, $f(\mathbf{L})$.

It involves computation of a lower triangular matrix $\mathbf{F}$. This is initialised to $\{\partial f(\mathbf{L})/\partial l_{ij}\}$. On completion of the backwards differentiation, $\mathbf{F}$ contains the derivatives of $f(\mathbf{L})$ with respect to the elements of $\mathbf{M}$. Smith (1995) states that the calculation of $\mathbf{F}$ (not including the work needed to compute $\mathbf{L}$) requires about twice as much work as one likelihood evaluation. Once $\mathbf{F}$ has been determined first derivatives of $f(\mathbf{L})$ can be obtained one at a time as $tr(\mathbf{F}\partial\mathbf{M}/\partial\theta_i)$, i.e., only one matrix $\mathbf{F}$ is required.

Meyer and Smith (1996) describe a REML algorithm utilising this technique to determine first and (observed) second derivatives of $\log\mathcal{L}$ for the case considered here. For $f(\mathbf{L}) = \log|\mathbf{C}| + \mathbf{y}'\mathbf{Py}$, the scalar is a function of the diagonal elements of $\mathbf{L}$ (see (12) and (13)). Hence, $\{\partial f(\mathbf{L})/\partial l_{ij}\}$ is a diagonal matrix with elements $2l_{ii}^{-1}$ for $i = 1,\ldots,M-1$ and $2l_{MM}$ in row M.

The non-zero derivatives of $\mathbf{M}$ have the same structure as the corresponding part (data *vs*. pedigree) part of $\mathbf{M}$.

$$\frac{\partial\mathbf{M}}{\partial\sigma_{A_{kl}}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\Sigma_A^{-1}\mathbf{D}_A^{\theta_i}\Sigma_A^{-1}\otimes\mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{17}$$

$$\frac{\partial\mathbf{M}}{\partial\sigma_{E_{kl}}} = \begin{bmatrix} \mathbf{X}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{X} & \mathbf{X}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{Z} & \mathbf{X}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{y} \\ \mathbf{Z}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{X} & \mathbf{Z}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{Z} & \mathbf{Z}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{y} \\ \mathbf{y}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{X} & \mathbf{y}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{Z} & \mathbf{y}'(\,\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}\,)\mathbf{y} \end{bmatrix} \tag{18}$$

As outlined above, $\mathbf{R}$ is blockdiagonal for animals. Hence, matrices $\partial\mathbf{R}^{-1}/\partial\sigma_{E_{kl}}$ have submatrices $-\Sigma_{E_w}^{-1}\mathbf{D}_w^{\theta_i}\Sigma_{E_w}^{-1}$, i.e., derivatives of $\mathbf{M}$ with respect to residual (co)variances can be set up in the same way as the 'data part' of $\mathbf{M}$.

The strategy outlined for the calculation of first derivatives of $\log\mathcal{L}$ does not require the inverse of the coefficient matrix $\mathbf{C}$. In contrast, Johnson and Thompson (1994 and 1995) and Gilmour *et al*. (1995) for the univariate case, and Madsen *et al*. (1994) and Jensen *et al*. (1995) for the multivariate case derive expressions for $\partial\log\mathcal{L}/\partial\theta_i$ based on (4), which require selected elements of $\mathbf{C}^{-1}$. Their scheme is computationally feasible due to the sparse matrix inversion method of Takahashi *et al*. (1973). Misztal (1994) claimed that each sparse matrix inversion took about two to three times as long as one likelihood evaluation, i.e. computational requirements for both alternatives to calculate first derivatives of $\log\mathcal{L}$ appear comparable.

Calculation of the average information

Define

$$\mathbf{b}_i = \frac{\partial\mathbf{V}}{\partial\theta_i}\mathbf{Py} \tag{19}$$

For $\theta_i = \sigma_{A_{kl}}$, $\partial\mathbf{V}/\partial\theta_i = \mathbf{Z}(\mathbf{D}_A^{\theta_i}\otimes\mathbf{A})\mathbf{Z}'$. This gives

$$\mathbf{b}_i = \mathbf{Z}(\mathbf{D}_A^{\theta_i}\Sigma_A^{-1}\otimes\mathbf{I}_{NA})\hat{\mathbf{a}} = (2-\delta_{kl})\sum_{m=1}^{t}(\sigma_A^{km}\mathbf{Z}_l + \sigma_A^{lm}\mathbf{Z}_k)\hat{\mathbf{a}}_m \tag{20}$$

where $\mathbf{I}_n$ is an identity matrix of size n, $\mathbf{Z}_m$ the the submatrix of $\mathbf{Z}$ and $\mathbf{a}_m$ the subvector of $\mathbf{a}$ for trait m, i.e. $\mathbf{b}_i$ is simply a weighted sum of solutions for animals in the data. For $\theta_i = \sigma_{E_{kl}}$ and $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{u}}$ the vector of residuals for (1) with subvectors $\hat{\mathbf{e}}_m$ for $m = 1,\ldots,t$,

$$\mathbf{b}_i = (\sum_{k=1}^{N}{}^{+}\mathbf{D}_{w_k}^{\theta_i}\Sigma_{E_{w_k}}^{-1})\hat{\mathbf{e}} = (2-\delta_{kl})\sum_{m=1}^{t}(\sigma_E^{km} + \sigma_E^{lm})\hat{\mathbf{e}}_m \tag{21}$$

Extension to models fitting additional random effects such as litter effects or maternal genetic effects is straightforward; see, for instance, Jensen *et al*. (1995) for corresponding expressions.

Using (19), (6) can be rewritten as

$$\mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_i}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{P}\mathbf{y} = \mathbf{b}'_i\mathbf{P}\mathbf{b}_j \tag{22}$$

Johnson and Thompson (1995) calculated vectors $\mathbf{P}\mathbf{b}_j$ as the residuals from repeatedly solving the mixed model equations pertaining to (1) with $\mathbf{y}$ replaced by $\mathbf{b}_j$ for $j = 1,\ldots,p$. On completion, (22) could be evaluated as simple vector products. Alternatively, define a matrix $\mathbf{B} = [\mathbf{b}_1 \mid \mathbf{b}_2 \mid \cdots \mid \mathbf{b}_p]$. Then consider the mixed model matrix with $\mathbf{y}$ replaced by $\mathbf{B}$, i.e., with the last row and column (for right hand sides) expanded to p rows and columns.

$$\mathbf{M}_B = \begin{bmatrix} \mathbf{C} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{B} \\ & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{B} \\ \mathbf{B}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{B}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{B}'\mathbf{R}^{-1}\mathbf{B} \end{bmatrix} \tag{23}$$

Factoring $\mathbf{M}_B$ or, equivalently, 'absorbing' $\mathbf{C}$ into the last p rows and columns of $\mathbf{M}_B$ then overwrites $\mathbf{B}'\mathbf{R}^{-1}\mathbf{B}$ with $\mathbf{B}'\mathbf{P}\mathbf{B}$ which has elements $\{\mathbf{b}'_i\mathbf{P}\mathbf{b}_j\}$ (Smith, 1994; pers. comm.). With the Cholesky factorisation of $\mathbf{C}$ already determined (to calculate $\log \mathcal{L}$), this is computationally undemanding.

EQUAL DESIGN MATRICES

For a simple animal model with all traits recorded at the same or corresponding times, design matrices are equal, i.e., (1) can be rewritten as

$$\mathbf{y} = (\mathbf{I}_t \otimes \mathbf{X}_0)\beta + (\mathbf{I}_t \otimes \mathbf{Z}_0)\mathbf{a} + \mathbf{e} \tag{24}$$

Meyer (1985) described a Method of Scoring algorithm for this case, exploiting a canonical transformation to reduce a t-variate analysis to t corresponding univariate analyses.

For Σ_E positive definite and Σ_A positive semi-definite, there exists a matrix $\mathbf{Q}$ such that Λ is a diagonal matrix with elements $\lambda_{ii} \geq 0$ which are the eigenvalues of $\Sigma_E^{-1}\Sigma_A$, and $\Omega = \mathbf{I}_t$ (e.g., Graybill, 1969).

$$\Lambda = \mathbf{Q}\Sigma_A\mathbf{Q}' \tag{25}$$
$$\Omega = \mathbf{Q}\Sigma_E\mathbf{Q}' \tag{26}$$

Transforming the data to

$$\mathbf{y}^* = (\mathbf{Q} \otimes \mathbf{I}_N)\mathbf{y} \tag{27}$$

then yields t new, 'canonical' traits which are uncorrelated and have unit residual variance. This makes the corresponding coefficient matrix in the MME blockdiagonal for traits, i.e.,

$$\mathbf{C}^* = \sum_{i=1}^{t} {}^+\mathbf{C}_i^* \qquad \text{with} \qquad \mathbf{C}_i^* = \begin{bmatrix} \mathbf{X}_0'\mathbf{X}_0 & \mathbf{X}_0'\mathbf{Z}_0 \\ \mathbf{Z}_0'\mathbf{X}_0 & \mathbf{Z}_0'\mathbf{Z}_0 + \lambda_{ii}^{-1}\mathbf{A}^{-1} \end{bmatrix} \tag{28}$$

Meyer (1991) described how the log likelihood (on the original scale) in this case can be computed trait by trait as the sum of univariate likelihoods on the canonical scale plus an adjustment for the transformation (last term in (29)).

$$\log \mathcal{L} = -\frac{1}{2}\left[N_A \sum_{i=1}^{t} \log \lambda_{ii} + \sum_{i=1}^{t} \log |\mathbf{C}_i^*| + \sum_{i=1}^{t} \mathbf{y}_i^{*'}\mathbf{P}_i^*\mathbf{y}_i^* + t \log |\mathbf{A}| \right.$$

$$\left. + (N - r(\mathbf{X}_0)) \log |\Sigma_E| \right] \tag{29}$$

with $\mathbf{y}_i^*$ the subvector of $\mathbf{y}^*$ for trait i and $\mathbf{P}_i^*$ the i-th diagonal block of the projection matrix on the canonical scale $\mathbf{P}^*$ which, like $\mathbf{C}^*$, is blockdiagonal for traits.

Terms required in (29) can be calculated by setting up and factoring, as described above, univariate MMM (on the canonical scale), $\mathbf{M}_i^*$, of size $M_0 = (M-1)/t+1$ each

$$\mathbf{M}_i^* = \begin{bmatrix} \mathbf{C}_i^* & \mathbf{r}_i^* \\ \mathbf{r}_i^{*'} & \mathbf{y}_i^{*'}\mathbf{y}_i^* \end{bmatrix} = \mathbf{L}_i^*\mathbf{L}_i^{*'} = \begin{bmatrix} \mathbf{L}_{C_i}^* & 0 \\ \mathbf{l}_i^{*'} & l_{MM_i}^* \end{bmatrix} \begin{bmatrix} \mathbf{L}_{C_i}^{*'} & \mathbf{l}_i^* \\ 0 & l_{MM_i}^* \end{bmatrix} \tag{30}$$

with $\mathbf{r}_i^* = (\mathbf{X}_0'\mathbf{y}_i^* \mid \mathbf{Z}_0'\mathbf{y}_i^*)'$.

Moreover, all first derivatives of $\log \mathcal{L}$ as well the average information matrix, both on the canonical scale, can be determined trait by trait.

First derivatives on the canonical scale

Consider the parameterisation of Meyer (1985) where $\boldsymbol{\theta}^*$, the vector of parameters on the canonical scale, has elements λ_{ij} and ω_{ij} for $i \leq j = 1, \ldots, t$, i.e. parameters are the (co)variances on the canonical scale.

The log likelihood on the canonical scale can be accumulated trait by trait, because Cholesky decompositions of individual MMM, $\mathbf{M}_i^*$, yield the submatrices and -vectors for trait i which are obtained when decomposing $\mathbf{M}^* = \mathbf{L}^*\mathbf{L}^{*'}$, i.e.,

$$\mathbf{L}^* = \begin{bmatrix} \mathbf{L}_C^* & \\ \mathbf{l}^{*'} & l_{MM}^* \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{C_1}^* & & & & \\ 0 & \mathbf{L}_{C_2}^* & & & \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & \mathbf{L}_{C_t}^* & \\ \mathbf{l}_1^{*'} & \mathbf{l}_2^{*'} & \cdots & \mathbf{l}_t^{*'} & l_{MM}^* \end{bmatrix} \tag{31}$$

On the original scale, $\mathbf{L}$ and $\mathbf{F}$ have the same sparsity structure (Smith, 1995). However, while $\lambda_{ij} = \omega_{ij} = 0$ for given Σ_A and Σ_E, the corresponding derivatives and estimates are not, unless the maximum of the likelihood has been attained. Hence, while the off-diagonal blocks of $\mathbf{L}_C^*$ are zero, the corresponding blocks of $\mathbf{F}^* = \partial f(\mathbf{L}^*)/\partial \mathbf{M}^*$ are not.

It can be shown that both the diagonal blocks of $\mathbf{F}^*$ corresponding to $\mathbf{L}_{C_i}^*$, $\mathbf{F}_{C_{ii}}^*$, and the row vectors corresponding to $\mathbf{l}_i^{*'}$, $\mathbf{f}_i^{*'}$, are identical to those obtained by backwards differentiation of $\mathbf{L}_i^*$. In other words, first derivatives with respect to the variance components on the canonical scale (λ_{ii} and ω_{ii}) can be obtained trait by trait from univariate analyses. Calculation of derivatives with respect to λ_{ij} and ω_{ij}, however, requires the off-diagonal blocks of $\mathbf{F}^*$ corresponding to traits i and j, $\mathbf{F}_{C_{ij}}^*$. Fortunately, as outlined in the Appendix, matrices $\mathbf{F}_{C_{ij}}^*$ can be determined indirectly from terms arising using the Cholesky decomposition and backwards differentiation for individual traits on the canonical scale.

From (17) and (18), first derivatives of $f(\mathbf{L}^*) = \log|\mathbf{C}^*| + \mathbf{y}^{*'}\mathbf{P}^*\mathbf{y}^*$ are then

$$\frac{\partial f(\mathbf{L}^*)}{\partial \lambda_{ij}} = -\lambda_{ii}^{-1}\lambda_{jj}^{-1} tr\left(\begin{bmatrix} 0 & 0 \\ 0 & \mathbf{A}^{-1} \end{bmatrix} \mathbf{F}_{C_{ij}}^*\right) \tag{32}$$

$$\frac{\partial f(\mathbf{L}^*)}{\partial \omega_{ij}} = -\left(tr\left(\begin{bmatrix} \mathbf{X}_0'\mathbf{X}_0 & \mathbf{X}_0'\mathbf{Z}_0 \\ \mathbf{Z}_0'\mathbf{X}_0 & \mathbf{Z}_0'\mathbf{Z}_0 \end{bmatrix} \mathbf{F}_{C_{ij}}^*\right) + \frac{1}{1+\delta_{ij}}\left(\mathbf{f}_i^{*'}\mathbf{r}_j^* + \mathbf{f}_j^{*'}\mathbf{r}_i^* + 2F_{MM}^*\mathbf{y}_i^{*'}\mathbf{y}_j^*\right)\right) \tag{33}$$

with F_{MM}^* the M−th diagonal element of $\mathbf{F}^*$. For $f(\mathbf{L}^*) = \log|\mathbf{C}^*| + \mathbf{y}^{*'}\mathbf{P}^*\mathbf{y}^*$, $F_{MM}^* = 1$.

Other terms required to determine the first derivatives on the canonical scale are

$$\frac{\partial \log|\mathbf{G}^*|}{\partial \theta_i^*} = \begin{cases} N_A\lambda_{kk}^{-1} & \text{for } \theta_i^* = \lambda_{kk} \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

$$\frac{\partial \log|\mathbf{R}^*|}{\partial \theta_i^*} = \begin{cases} N & \text{for } \theta_i^* = \omega_{kk} \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

7

where $\mathbf{G}^*$ and $\mathbf{R}^*$ are the canonical scale equivalents to $\mathbf{G}$ and $\mathbf{R}$, respectively.

Average information on the canonical scale

For $\mathbf{G}^* = \Lambda \otimes \mathbf{A}$, $\mathbf{R}^* = \mathbf{I}_{tN}$, and thus $\mathbf{V}^* = Var(\mathbf{y}^*)$ blockdiagonal for traits, (20) and (21) simplify to

$$\mathbf{b}_i^* = \frac{\partial \mathbf{V}^*}{\partial \lambda_{kl}} \mathbf{P}^* \mathbf{y}^* = (\mathbf{D}_{kl} \otimes \mathbf{I}_N) \left\{ \lambda_j^{-1} \mathbf{Z}_0 \hat{\mathbf{a}}_j^* \right\} \tag{36}$$

and

$$\mathbf{b}_i^* = \frac{\partial \mathbf{V}^*}{\partial \omega_{kl}} \mathbf{P}^* \mathbf{y}^* = (\mathbf{D}_{kl} \otimes \mathbf{I}_N) \left\{ \hat{\mathbf{e}}_j^* \right\} \tag{37}$$

i.e., vectors $\mathbf{b}_i^*$ are zero except for subvectors for traits k and l, $\mathbf{b}_{ik}^* = \mathbf{s}_l$ and $\mathbf{b}_{il}^* = \mathbf{s}_k$, with $\mathbf{s}_j$ standing in turn for $\lambda_{jj}^{-1} \mathbf{Z}_0 \hat{\mathbf{a}}_j^*$ and $\hat{\mathbf{e}}_j^*$.

With $\mathbf{P}^*$ blockdiagonal for traits, $\theta_i^* = \lambda_{kl}$ or ω_{kl} ($k \leq l$) and $\theta_j^* = \lambda_{mn}$ or ω_{mn} ($m \leq n$), this gives

$$\mathbf{b}_i^{*'} \mathbf{P}^* \mathbf{b}_j^* = \begin{cases} 0 & \text{for } k \neq m, l \neq n \\ \mathbf{b}_{ik}^{*'} \mathbf{P}_m^* \mathbf{b}_{jm}^* + \mathbf{b}_{il}^{*'} \mathbf{P}_n^* \mathbf{b}_{jn}^* = \mathbf{s}_l' \mathbf{P}_m^* \mathbf{s}_n + \mathbf{s}_k' \mathbf{P}_n^* \mathbf{s}_m & \text{for } k = m \neq l = n \\ \mathbf{b}_{ik}^{*'} \mathbf{P}_m^* \mathbf{b}_{jm}^* = \mathbf{s}_l' \mathbf{P}_m^* \mathbf{s}_n & \text{otherwise} \end{cases} \tag{38}$$

Hence, calculation of the average information on the canonical scale requires all terms $\mathbf{s}_i' \mathbf{P}_k^* \mathbf{s}_j$ for $i \leq j = 1, \ldots, 2t$ and $k = 1, t$. These can be obtained trait by trait, analogously to the procedure described above for the general case. After the Cholesky factorisation of $\mathbf{M}_k^*$ has been carried out, solution $\mathbf{a}_k^*$ for animals' additive genetic effects and residuals $\mathbf{e}_k^*$ for trait k are obtained, storing the Cholesky factor $\mathbf{L}_k^*$. Define a matrix $\mathbf{S}$ of size $N \times 2t$ with columns equal to vectors $\mathbf{s}_i$. $\mathbf{S}$ is the canonical scale equivalent to $\mathbf{B}$ above. Once all columns of $\mathbf{S}$ have been evaluated, set up a matrix

$$\mathbf{M}_{S_k}^* = \begin{bmatrix} & \mathbf{C}_k^* & \mathbf{X}_0' \mathbf{S} \\ & & \mathbf{Z}_0' \mathbf{S} \\ \mathbf{S}' \mathbf{X}_0 & \mathbf{S}' \mathbf{Z}_0 & \mathbf{S}' \mathbf{S} \end{bmatrix} \tag{39}$$

for each trait. This is the MMM for trait k on the canonical scale with $\mathbf{y}_k^*$ replaced by $\mathbf{S}$. The matrix $\mathbf{S'S}$ has elements $\mathbf{s}_i' \mathbf{s}_j$ which are the sum of squares and crossproducts off the vectors of (weighted) solutions and residuals. Absorbing $\mathbf{C}_k^*$ into $\mathbf{S'S}$ (using the stored matrix $\mathbf{L}_k^*$) then overwrites this matrix with $\mathbf{S'P}_k^* \mathbf{S}$ which has elements $\mathbf{s}_i' \mathbf{P}_k^* \mathbf{s}_j$. After all t traits have been processed $\mathbf{b}_i^{*'} \mathbf{P}^* \mathbf{b}_j^*$, (twice) the average information, can be 'assembled' according to (38).

Derivatives on the original scale

Let $\mathbf{H}^*$, with elements $\frac{1}{2} \mathbf{b}_i^{*'} \mathbf{P}^* \mathbf{b}_j^*$, and $\mathbf{g}^*$, with elements $\partial \log \mathcal{L}^* / \partial \theta_i^*$, denote the average information matrix and vector of gradients on the canonical scale, respectively. Corresponding terms on the original scale are then

$$\mathbf{H} = \mathbf{J} \mathbf{H}^* \mathbf{J}' \tag{40}$$
$$\mathbf{g} = \mathbf{J} \mathbf{g}^* \tag{41}$$

where $\mathbf{J}$ with elements $\partial \theta_i^* / \partial \theta_j$ is the Jacobian matrix of θ with respect to θ^*. From (25) and (26), $\mathbf{J}$ has non-zero elements

$$\frac{\partial \lambda_{ij}}{\partial \sigma_{A_{mn}}} = \frac{1}{1 + \delta_{mn}} (q_{im} q_{jn} + q_{in} q_{jm}) \tag{42}$$

$$\frac{\partial \omega_{ij}}{\partial \sigma_{E_{mn}}} = \frac{1}{1 + \delta_{mn}} (q_{im} q_{jn} + q_{in} q_{jm}) \tag{43}$$

A numerical example illustrating calculations is given in the Appendix.

Alternative parameterisation

The above parameterisation requires switching between the original and canonical scale (or accumulating the transformations). Alternatively, as done by Meyer (1991) for a derivative-free algorithm, θ^* can be defined to have t elements λ_{ii} ($i = 1, \ldots, t$) and t^2 elements q_{ij} ($i, j = 1, \ldots, t$), i.e. the 'genetic' variances on the canonical scale and the elements of the transformation matrix $\mathbf{Q}$. This allows estimation to be carried out on the canonical scale. Moreover, as outlined by Meyer (1991) for a derivative-free algorithm, evaluation of $\log \mathcal{L}$ for given λ_{ii} requires scalar calculations involving the elements of $\mathbf{Q}$ only, i.e. maximising the conditional $\log \mathcal{L}$ (for given λ_{ii}) with respect to q_{ij} is computationally inexpensive. This allowed the maximum of $\log \mathcal{L}$ to be located in a two-step procedure with computational requirements equivalent to those of a t parameter (rather than $t(t+1)$) analysis.

While derivatives with respect to q_{ij} (not shown) require the off-diagonal blocks of $\mathbf{F}$, derivatives with respect to λ_{ii} do not (see (32)). Hence a nested maximisation, using an average information step to estimate parameters λ_{ii} and a derivative-free procedure to maximize $\mathcal{L}$ with respect to q_{ij} for given λ_{ii} appears computationally advantageous in this case.

REDUCED RANK COVARIANCE MATRICES

Forcing the estimate of the genetic covariance matrix $\hat{\Sigma}_A$ to have reduced rank ($k_A < t$), results in a number of zero eigenvalues on the canonical scale. In this case, several terms contributing to $\log \mathcal{L}$ or its derivatives are constant or depend on the canonical transformation ($\mathbf{Q}$) only. Thus they need to be evaluated only once per analysis, effectively reducing the computational requirements per round of iteration to those for univariate analyses for the k_A non-zero eigenvalues.

Likelihood For $\lambda_{ii} \to 0$, contributions to $\log \mathcal{L}$ (29) become

$$N_A \log \lambda_{ii} + \log |\mathbf{C}_i^*| \to \log |\mathbf{X}_0' \mathbf{X}_0| + \log |\mathbf{A}^{-1}| \tag{44}$$

and

$$\mathbf{y}_i^{*'} \mathbf{P}_i^* \mathbf{y}_i^* \to \sum_{m=1}^{t} \sum_{n=1}^{t} q_{im} q_{jn} \mathbf{y}_m' \left(\mathbf{I} - \mathbf{X}_0 (\mathbf{X}_0' \mathbf{X}_0)^- \mathbf{X}_0' \right) \mathbf{y}_n \tag{45}$$

While the former is a constant, determined by the data structure only, the latter depends on the canonical transformation. However, as (45) shows, it can be calculated for any $\mathbf{Q}$ from the corresponding residual sums of squares (SS) and crossproducts (CP) on the original scale. Let $\mathbf{Y}$ denote a matrix of size $N \times t$ with columns equal to vectors of observations $\mathbf{y}_i$. Both $\log |\mathbf{X}_0' \mathbf{X}_0|$ and quadratics $\mathbf{y}_m' (\mathbf{I} - \mathbf{X}_0 (\mathbf{X}_0' \mathbf{X}_0)^- \mathbf{X}_0') \mathbf{y}_n$ can then be determined by factoring the matrix

$$\mathbf{M}_X = \left[\begin{array}{cc} \mathbf{X}_0' \mathbf{X}_0 & \mathbf{X}_0' \mathbf{Y} \\ \mathbf{Y}' \mathbf{X}_0 & \mathbf{Y}' \mathbf{Y} \end{array} \right] \tag{46}$$

First derivatives To estimate genetic and residual covariance components, only the derivatives on the canonical scale (with respect to λ_{ij} or ω_{ij}) are required for which the eigenvalues for both canonical traits (λ_{ii} and λ_{jj}) are non-zero. This reduces the number of Cholesky decompositions (of matrices $\mathbf{M}_i^*$) and corresponding backwards differentiations (to obtain matrices $\mathbf{F}_{C_{ii}}^*$ and vectors $\mathbf{f}_i^*$) to be carried out to the number of non-zero eigenvalues, k_A. Moreover, only $k_A(k_A - 1)/2$ instead of $t(t-1)/2$ off-diagonal blocks $\mathbf{F}_{C_{ij}}^*$ need to be evaluated. This can reduce computational requirements per round of iteration dramatically.

Average information As shown above, the average information matrix can be constructed from residual SS and CP in the vectors of random effects solutions and residuals. For $\lambda_{kk} = 0$, $\hat{\mathbf{a}}_k^* = 0$. Residuals on the canonical scale are then a linear combination of residuals on the original scale, $\hat{\mathbf{e}}_k^* = \sum_{m=1}^{t} q_{km} \hat{\mathbf{e}}_m$ with $\hat{\mathbf{e}}_m = \mathbf{y}_m - \mathbf{X}_0 \hat{\beta}_m$. Again the

latter need to be evaluated only once. Terms required for the AI can then be obtained as before, replacing $\mathbf{M}^*_{S_k}$ in (39) with

$$
\begin{bmatrix}
\mathbf{X}'_0\mathbf{X}_0 & \mathbf{X}'_0\mathbf{S} \\
\mathbf{S}'\mathbf{X}_0 & \mathbf{S}'\mathbf{S}
\end{bmatrix}
\tag{47}
$$

if $\lambda_{kk} = 0$. For multiple eigenvalues equal to zero, (47) only needs to be evaluated once (per iterate).

COVARIANCE FUNCTIONS

The 'infinitely-dimensional' model

Consider 'repeated' measurements taken at a number of ages (or equivalent), with potentially infinitely many records. The covariance between records taken at ages l and m can be expressed as

$$
\mathcal{F}(a_l, a_m) = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \phi_i(a_l)\phi_j(a_m)K_{ij}
\tag{48}
$$

where $\mathcal{F}$ is the covariance function and $\mathbf{K}$ with elements K_{ij} is the pertaining matrix of coefficients, and a_m is the $m-$th age, standardised to the interval for which the polynomials are defined. Φ is a matrix of orthogonal polynomial functions evaluated at the given ages with elements $\phi_{ij} = \phi_j(a_i)$, the $j-$th polynomial evaluated for age i, and k denotes the order of polynomial fit. Conceptually $k = \infty$, but in practice $k \leq t$ for observations at t ages. Kirkpatrick *et al.* (1990 and 1994) suggested the use of the so-called Legendre polynomials (see Abramowitz and Stegun, 1965) which span the interval from -1 to 1.

Note that polynomials include a scalar term, i.e. for t records a full order fit involves terms to the power $0, \ldots, t-1$. From (48), a covariance matrix can be rewritten as

$$
\Sigma = \Phi\mathbf{K}\Phi'
\tag{49}
$$

For a reduced order fit, $k < t$, Φ is rectangular with k columns and only $k(k+1)/2$ coefficients K_{ij} need to be estimated.

Let $\Sigma_A = \Phi_A\mathbf{K}_A\Phi'_A$ for an order of fit k_A, and $\mathbf{K}_A$ the matrix of coefficients for the corresponding covariance function $\mathcal{A}$. Further, partition the error covariance matrix into components due to permanent and temporary environmental effects, $\Sigma_E = \Sigma_R + \Sigma_\varepsilon$. Under the 'finite model', these usually cannot be disentangled unless there are repeated records for the same age. Assume the latter represent independent 'measurement errors', i.e. $\Sigma_\varepsilon = \text{Diag}\left\{\sigma^2_{\varepsilon_i}\right\}$, and that the former can be described by a CF, $\mathcal{R}$, i.e. $\Sigma_R = \Phi_R\mathbf{K}_R\Phi'_R$ with order of fit k_R. Fitting measurement errors separately together with $\mathcal{R}$ to the order $t-1$ yields an equivalent model to a full order fit for Σ_E. Hence the maximum for k_R is $t-1$ rather than t.

General case

Estimates of the elements of the coefficient matrices of the covariance functions (and the measurement errors) can be estimated by REML using algorithms for the multivariate estimation of covariance components together with a simple reparameterisation.

Likelihood   As outlined by Meyer and Hill (1997), $\log \mathcal{L}$ (9) can be rewritten as

$$
\begin{aligned}
\log \mathcal{L} = & -\frac{1}{2}\left[const + \log|\mathbf{C}| + \mathbf{y}'\mathbf{P}\mathbf{y} + t(\log|\mathbf{A}| + \log|\mathbf{D}|)\right. \\
& \left. + N_A(\log|\mathbf{K}_A| + \log|\Phi_A\Phi'_A|) + N\log\left|\Phi_R\mathbf{K}_R\Phi'_R + \text{Diag}\left\{\sigma^2_{\varepsilon_i}\right\}\right|\right]
\end{aligned}
\tag{50}
$$

Under the CF model, the vector of parameters to be estimated is η with elements $K_{A_{ij}}$ for $i \leq j = 1, \ldots k_A$, $K_{R_{ij}}$ for $i \leq j = 1, \ldots k_R$ and $\sigma^2_{\varepsilon_i}$ for $i = 1, \ldots, t$. For r CFs to be estimated, it has minimum length $r + t$ (fitting all CFs to order 1 and assuming all $\sigma^2_{\varepsilon_i}$ to be distinct) and maximum length $rt(t+1)/2$ (fitting all CFs to full order).

10

Derivatives Note that the elements of matrices Φ only depend on the ages at which records were taken and the polynomial function chosen. For any vector η a corresponding vector of covariance components (θ) can be calculated using (49). First and 'average' second derivatives of $\log \mathcal{L}$ can thus be determined with respect to the elements of θ as described above (section on estimation of covariance components in the general cases) and then be transformed to the 'covariance function scale', using

$$\frac{\partial \log \mathcal{L}}{\partial \eta} = \mathbf{J}\frac{\partial \log \mathcal{L}}{\partial \theta} \tag{51}$$

$$\frac{\partial^2 \mathbf{y'Py}}{\partial \eta \partial \eta'} = \mathbf{J}\frac{\partial^2 \mathbf{y'Py}}{\partial \theta \partial \theta'}\mathbf{J'} \tag{52}$$

where $\mathbf{J}$ with elements $\partial \theta_i/\partial \eta_j$ is the Jacobian matrix of θ with respect to η. From (48), non-zero elements of $\mathbf{J}$ are

$$J_{ij} = \frac{1}{1+\delta_{rs}}(\phi_{mr}\phi_{ns} + \phi_{ms}\phi_{nr}) \tag{53}$$

for $\eta_i = K_{A_{mn}}$ and $\theta_j = \sigma_{A_{rs}}$, or $\eta_i = K_{R_{mn}}$ and $\theta_j = \sigma_{E_{rs}}$, and

$$J_{ij} = 1 \tag{54}$$

for $\eta_i = \sigma_{\varepsilon_m}^2$ and $\theta_j = \sigma_{E_{mm}}$.

Note that a reduced order fit for $\mathcal{A}$ ($k_A < t$) implies a genetic covariance matrix Σ_A of reduced rank. This may lead to computational problems when applying 'standard' methodology to factor the MMM (such as the Cholesky decomposition and its automatic differentiation) as described above. For practical applications, this can be overcome by setting the $t - k_A$ zero eigenvalues to an operational zero, i.e. a small non-zero value such as 10^{-5} or 10^{-6}.

As for the estimation of covariance components, extensions to other models including additional random effects are straightforward.

Equal design matrices

Assuming measurement errors are greater than zero, $\Sigma_E = \Sigma_R + \Sigma_\varepsilon$ is positive definite, regardless of the order of fit for $\mathcal{R}$. Hence, the transformation to canonical scale exists and the log likelihood can be determined as for the 'finite model', summing terms from univariate analyses on the canonical scale as described above.

For the estimation of covariance components, the canonical transformation doubled as a tool to reduce computational requirements (allowing calculations to be carried out for one trait at a time) and as a reparameterisation, i.e. the likelihood was maximised with respect to the eigenvalues and elements of eigenvectors of the canonical decomposition of Σ_A and Σ_E. For the CF model there are 3 matrices to be considered, $\mathbf{K}_A$ and $\mathbf{K}_R$, or Σ_A and Σ_R derived from them, and Σ_ε. Decompositions diagonalising several matrices exist. Lin and Smith (1990), for instance, used the common principal components algorithm of Flury and Constantine (1985) as an equivalent to the canonical decomposition for a multivariate mixed model with several random effects. However, this required the matrices to be transformed to be positive definite, and is thus not suitable for this application.

Hence only the first parameterisation described above, namely θ^* having elements λ_{ij} and ω_{ij} for $i \le j = 1,\ldots,t$ is suitable for the estimation of CFs. First derivatives and and average information on the canonical scale can then be determined and transformed back to the original scale as described above. The Jacobian $\mathbf{J}$ in this case has non-zero elements

$$\frac{\partial \lambda_{ij}}{\partial K_{A_{mn}}} = \frac{1}{1+\delta_{mn}}(\psi_{im}\psi_{jn} + \psi_{in}\psi_{jm}) \tag{55}$$

$$\frac{\partial \omega_{ij}}{\partial K_{R_{mn}}} = \frac{1}{1+\delta_{mn}}(\psi_{im}\psi_{jn} + \psi_{in}\psi_{jm}) \tag{56}$$

$$\frac{\partial \omega_{ij}}{\partial \sigma_{\varepsilon_m^2}} = q_{im}q_{jm} \tag{57}$$

where ψ_{mn} are the elements of $\Psi = \mathbf{Q}\Theta$.

DISCUSSION

A strategy has been outlined for computing the log likelihood in a multivariate mixed model, together with its first and 'average' second derivatives with respect to covariance components or the coefficients of covariance functions. These can be used a (modified) Newton-type estimation procedure,

$$\theta^{t+1} = \theta^t - (\mathbf{H}^t + \kappa^t \mathbf{I})^{-1} \mathbf{g}^t \tag{58}$$

(Marquard, 1963), together with an additional transformation to ensure estimates to be within the parameter space; see Meyer and Smith (1996) for details.

For a model with only one random factor and equal design matrices for all traits, calculations can be carried out trait by trait, reducing computational requirements to those of a series of univariate analyses. This is feasible through a canonical decomposition of the genetic and residual covariance matrices and a corresponding, linear transformation of the data to new, uncorrelated variables.

In that case, fitting a covariance function to less than full order or forcing estimated genetic covariance matrices to be of reduced rank is equivalent to fixing eigenvalues on the canonical scale at zero. Most terms pertaining to zero eigenvalues then only need to be calculated once per analysis, resulting in considerable reductions of computational requirements. In essence, it reduces work required in each round of an iterative solution scheme to that equivalent to k_A univariate analyses. This is particularly useful for a comparatively large number of highly correlated measurements where a covariance function of low order suffices to describe the data adequately. In contrast, for analyses where a transformation to canonical scale is not feasible, the complete t-variate MMM needs to be set up, factored and differentiated in each round of iteration, even if CFs are only fitted to order k_A.

Making computational demands proportional to the order of fit for the genetic CF (or matrix) encourages an 'upwards' strategy : increasing the order of fit one by one allows a likelihood ratio test to be performed at each step. The minimum number of parameters describing the data is found when the likelihood ceases to increase significantly when the order of fit is increased. It is envisioned that estimation of reduced order covariance matrices or functions will become the standard procedure for high-dimensional multivariate analyses.

ACKNOWLEDGEMENTS

REFERENCES

Abramowitz M, Stegun IA (1965) *Handbook of Mathematical Functions*. Dover, New York

Flury B, Constantine G (1985) Algoritm AS 211 : The F-G diagonalization algorithm. *Appl Stat* 177–180,

George A, Liu J W-H (1981) *Computer Solution of Large Positive Definite Systems*. Prentice-Hall Inc, Englewood Cliffs, New Jersey 07632

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML : An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440–1450

Graser H-U, 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

Graybill, FA (1969) *Introduction to Matrices with Applications in Statistics*. Wadsworth Publishing Company, Inc, Belmont, California

Harville, DA (1977) Maximum likelihood approaches to variance component estimation and related problems. *J Amer Stat Ass* 72, 320–338

Jensen J, Mao IL (1988) Transformation algorithms in analysis of single trait and of multitrait models with equal design matrices and one random factor per trait : a review. *J Anim Sci* 66, 2750–2761

Jensen J, Mantysaari E, Madsen P, Thompson R (1995) Restricted Maximum Likelihood estimation of (co)variance components in multivariate linear models using average of observed and expected information. 2nd European Workshop on Advanced Biometrical Methods in Animal Breeding, Salzburg, Austria, June 12–20, 1995, Mimeo, 17p.

Johnson DL, Thompson R (1994) Restricted Maximum Likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *Proc*. 5th *World Congr Genet Appl Livest Prod, University of Guelph, Guelph, Canada* 18, 410–413

Johnson DL, Thompson R (1995) Restricted Maximum Likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J Dairy Sci* 78, 449–456

Kirkpatrick M, Heckman N (1989) A quantitative genetic model for growth, shape and other infinite-dimensional characters. *J Math Biol* 27, 429–450

Kirkpatrick M, Lofsvold D, Bulmer M (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979-993

Kirkpatrick M, Hill WG, Thompson R (1994) Estimating the covariance structure of traits during growth and aging, illustrated with lactations in dairy cattle. *Genet Res* 64, 57–69

Lin CY, Smith S.P. (1990) Transformation of multi-trait to unitrait mixed model analysis of data with multiple random effects. *J Dairy Sci* 73, 2494–2502

Lindstrom MJ, Bates DM (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J Amer Stat Ass* 83, 1014–1022

Madsen P, Jensen J, Thompson R (1994) Estimation of (co)variance components by REML in multivariate mixed linear models using average of observed and expected information. *Proc*. 5th *World Congr Genet Appl Livest Prod, University of Guelph, Guelph, Canada* Vol 22, 19–22

Marquardt DW (1963) An algorithm for least squares estimation of nonlinear parameters. *SIAM J* 11, 431–441

Meyer K (1985) Maximum Likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41, 153–166

Meyer K (1989) Restricted Maximum Likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet Select Evol* 21, 317–340

Meyer K (1991) Estimating variances and covariances for multivariate Animal Models by Restricted Maximum Likelihood. *Genet Select Evol* 23, 67–83

Meyer K, Smith SP (1996) Restricted Maximum Likelihood estimation for animal models using derivatives of the likelihood. *Genet Select Evol* 28, 23–49

Meyer K, Hill WG (1997) Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by Restricted Maximum Likelihood. *Livest Prod Sci* , (in press)

Misztal I (1994) Comparison of computing properties of derivative and derivative-free algorithms in variance component estimation by REML. *J Anim Breed Genet* 111, 346–355

Searle SR (1982) *Matrix Algebra Useful for Statistics*. John Wiley & Sons, New York

Smith SP (1995) Differentiation of the Cholesky algorithm. *J Comp Graph Stat* 4, 134–147

Takahashi K, Fagan J, Chin MS (1973) Formation of sparse bus impedance matrix and its application to short circuit study. Proc 8th Inst PICA Conf Minneapolis, MN, p63-69; cited Smith (1995)

Appendix

Calculation of off-diagonal blocks of F

Smith (1995) gives pseudo-code for backwards differentiation of the Cholesky factor of a symmetric matrix. This can be adapted to calculating a selected submatrix only for the case where the corresponding part of the Cholesky factor has zero elements.

Calculating the off-diagonal block of $\mathbf{F}^*$ for traits i and j, $\mathbf{F}^*_{C_{ij}}$ indirectly, requires components from univariate analyses on the canonical scale, namely $\mathbf{L}^*_{C_i}$ with elements L^i_{kl} and $\mathbf{l}^*_i$ with elements l^i_k, the parts of the Cholesky factor of $\mathbf{M}^*_i$ corresponding to the coefficient matrix and vector of right hand sides, respectively, and $\mathbf{f}^*_j$ with elements f^j_k, the 'F matrix' analogue to the vector of right hand sides for trait j. $\mathbf{F}^*_{C_{ij}}$ with elements F_{kl} can then be calculated as follows :

1. Initialise $\mathbf{F}^*_{C_{ij}}$ to $\mathbf{f}^*_j \mathbf{l}^{*'}_i$, i.e. for $k, l = 1, \ldots, M_0 - 1$, $F_{kl} := f^j_k l^i_m$

2. Calculate columns of $\mathbf{F}^*_{C_{ij}}$ one at a time, starting with the last column, i.e. for $k = M_0 - 1, \ldots, 1$

   (a) Adjust for columns already evaluated, i.e. for $m = M_0 - 1, \ldots, k+1$ and $l = 1, M_0 - 1$, $F_{lk} := F_{lk} + F_{lm} L^i_{mk}$
   (b) Divide by the pivot, i.e. for $l = 1, M_0 - 1$, $F_{lk} := -F_{lk}/L^i_{kk}$

where ':=' indicates that the left hand side of the equation is overwritten with the quantity on the right.

As noted by Smith (1995), only selected elements of $\mathbf{F}$ (in the general case) need to be evaluated - adjustments only occurring for non-zero elements of the Cholesky factor. Hence $\mathbf{F}$ has the same sparsity structure as $\mathbf{L}$. Similar considerations hold in this case, reducing computational requirements in practical applications substantially. For instance, columns of $\mathbf{F}^*_{C_{ij}}$ are not required for adjustments to other columns if the corresponding row of $\mathbf{L}^*_i$ has no non-zero off-diagonal elements. Hence in these columns only elements corresponding to potentially non-zero elements in the derivatives of $\mathbf{M}^*$ need to be evaluated. Other redundancies could be perceived. For a particular analysis it might be worth carrying out a symbolic factorisation of the MMM for all traits on the original scale to determine the sparsity structure of the 'full $\mathbf{L}$' and thus the minimum number of elements of $\mathbf{F}^*_{C_{ij}}$ which need to be evaluated. This would have to be carried out only once per analyses prior to the iterative estimation scheme.

Numerical Example

Consider the example given by Meyer (1991), consisting of two traits measured on 284 mice. Table 1 summarises starting values, intermediate results for the first iterate and estimates over rounds of iterations. Convergence is reached rapidly even though starting values for covariances were far from the eventual estimates, the AI algorithm performing similar to an algorithm using observed or expected second derivatives of log $\mathcal{L}$ (see Meyer and Smith, 1996).

For the starting values given the canonical transformation is

$$\mathbf{Q} = \begin{bmatrix} 0.70840 & -0.07868 \\ -0.23025 & 0.31833 \end{bmatrix}$$

with eigenvalues $\lambda_{11} = 1.96406$ and $\lambda_{22} = 0.50389$. Quadratics on the canonical scale $(\mathbf{y}^{*'}_i \mathbf{P}^*_i \mathbf{y}^*_i)$ are 314.77078 and 375.36955 for the two traits and the corresponding log $|\mathbf{C}^*_i|$ are 241.85744 and 554.08056, respectively. With 329 animals in the an analysis and $r(\mathbf{X}_0) = 2$, this gives log $\mathcal{L}$ omitting the term $t \log |\mathbf{A}|$, of -1172.3761 (c.f. (29)).

First derivatives of log $|\mathbf{G}^*|$ and log $|\mathbf{R}^*|$ calculated according to (34) and (35) are given in Table 1. Terms $\mathbf{y}^{*'}_i \mathbf{y}^*_j$ ($i \leq j = 1, 2$) and $\mathbf{f}^{*'}_i \mathbf{r}^*_j$ ($i, j = 1, 2$) are 42285.252, 50449.049 and 64091.809, and -83940.962, -101237.320,

14

Table 1 : First derivatives of the log likelihood ($\mathcal{L}$) and average information together with estimates over rounds of iteration for numerical example.

$i =$		Parameter					log $\mathcal{L}$
	1	2	3	4	5	6	
Starting values							
θ_i	4.7000	4.0000	8.3000	2.5000	3.0000	12.9000	-1172.3761
θ_i^*	1.9641	0	0.5039	1	0	1	
First derivatives							
$\partial \log |\mathbf{G}^*|/\partial\theta_i^*$	167.51	0	652.92	0	0	0	
$\partial \log |\mathbf{R}^*|/\partial\theta_i^*$	0	0	0	284	0	284	
$\partial f(\mathbf{L}^*)/\partial\theta_i^*$	-178.270	101.222	-719.163	-303.636	140.415	-351.989	
$\partial \log \mathcal{L}/\partial\theta_i^*$	5.3802	-50.6111	33.1228	9.8182	-70.2077	33.9945	
$\partial \log \mathcal{L}/\partial\theta_i$	12.7110	-17.7853	4.6575	18.1806	-23.1819	5.2641	
Average information[a]							
	-11.9738	8.7052	0.0000	-17.3585	8.2079	0.0000	
		-56.4919	21.2427	8.2079	-62.6270	14.6015	
	-6.5215		-75.0270	0.0000	14.6015	-58.8969	
	5.5025	-7.1041		-43.0093	18.9830	0.0000	
	-1.1686	2.1074	-0.9169		-134.6006	27.7462	
	-7.7992	5.6077	-0.9984	-18.3064		-109.2794	
	5.6077	-6.6871	1.7184	12.2914	-13.8844		
	-0.9984	1.7184	-0.7214	-2.0238	3.3428	-1.3550	
Estimates over rounds of iteration							
1	4.0476	2.5948	5.8965	3.0589	0.6548	15.0535	-1148.2602
2	4.2164	1.8607	8.1262	2.9166	0.9662	14.0864	-1144.3823
3	5.0011	0.0370	8.1894	2.6981	2.0881	13.0165	-1142.5610
4	4.4354	0.1713	7.8996	2.6046	2.0547	13.1032	-1142.0188
5	4.3643	0.1484	7.9252	2.6225	2.0691	13.0839	-1142.0164
6	4.3764	0.1495	7.9148	2.6158	2.0688	13.0897	-1142.0163

[a]Upper triangle : canonical scale, lower triangle : variance component scale

-101089.518 and -127432.879, respectively. Application of (32) and (33) then gives the first derivatives (canonical scale) of $f(\mathbf{L}^*)$, with corresponding derivatives $\partial \log \mathcal{L}/\partial\theta_i^* = -1/2\,(\partial \log |\mathbf{G}^*|/\partial\theta_i^* + \partial \log |\mathbf{R}^*|/\partial\theta_i^* + \partial f(\mathbf{L}^*)/\partial\theta_i^*)$ as shown in Table 1. From (42) and (43), the Jacobian is

$$\mathbf{J} = \begin{bmatrix} 0.50182 & -0.16311 & 0.05301 & 0 & 0 & 0 \\ -0.11147 & 0.24362 & -0.14659 & 0 & 0 & 0 \\ 0.00619 & -0.02505 & 0.10134 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.50182 & -0.16311 & 0.05301 \\ 0 & 0 & 0 & -0.11147 & 0.24362 & -0.14659 \\ 0 & 0 & 0 & 0.00619 & -0.02505 & 0.10134 \end{bmatrix}$$

and (41) yields derivatives on the original scale as shown (Table 1). The average information on the canonical scale (upper triangle in Table 1) is calculated according to (35) to (38), and (40) gives the corresponding values on the original scale.