Original article

Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices

Karin MEYER*, Mark KIRKPATRICK

Animal Genetics and Breeding Unit**, University of New England,
Armidale NSW 2351, Australia
Section of Integrative Biology, 1 University Station C-0930, University of Texas,
Austin, Texas 78712, USA

Abstract – Principal component analysis is a widely used 'dimension reduction' technique, albeit generally at a phenotypic level. It is shown that we can estimate genetic principal components directly through a simple reparameterisation of the usual linear, mixed model. This is applicable to any analysis fitting multiple, correlated genetic effects, whether effects for individual traits or sets of random regression coefficients to model trajectories. Depending on the magnitude of genetic correlation, a subset of the principal component generally suffices to capture the bulk of genetic variation. Corresponding estimates of genetic covariance matrices are more parsimonious, have reduced rank and are smoothed, with the number of parameters required to model the dispersion structure reduced from $k(k + 1)/2$ to $m(2k − m + 1)/2$ for k effects and m principal components. Estimation of these parameters, the largest eigenvalues and pertaining eigenvectors of the genetic covariance matrix, *via* restricted maximum likelihood using derivatives of the likelihood, is described. It is shown that reduced rank estimation can reduce computational requirements of multivariate analyses substantially. An application to the analysis of eight traits recorded *via* live ultrasound scanning of beef cattle is given.

covariances / principal components / restricted maximum likelihood / reduced rank

1. INTRODUCTION

In most quantitative genetic analyses we are interested in multiple, correlated genetic effects per individual. Predominantly, these are additive genetic effects for different traits, but recently applications which model trajectories for traits measured repeatedly per individual through sets of correlated, (genetic) random regression (RR) coefficients, have found increasing use. Traditionally, the corresponding genetic covariance matrices have been considered

---

unstructured, *i.e.* for *k* traits (or RR coefficients) there were $k(k + 1)/2$ distinct (co)variance components to be estimated. By and large, restrictions on estimates were imposed only to ensure that estimates were within the parameter space, *i.e.* that all variances and conditional variances were positive, that all correlation estimates were in the range of −1 to +1, and that all partial correlations were consistent with each other. In statistical terms, this was equivalent to the requirement that the estimated covariance matrix was positive semi-definite, *i.e.* that none of its eigenvalues was negative.

In contrast, other areas of statistics have long since assumed and estimated structured covariance matrices. On the one hand, these have been associated with longitudinal or spatial data, or simply 'repeated' records. Jennrich and Schluchter [21] and Wolfinger [57], for instance, discuss a number of common structures, including compound symmetry, autoregressive and factor-analytic structure, and random coefficient models (or see books by Diggle *et al.* [5] and Lindsey [32]). On the other hand, principal component (PC) analysis, and the closely related factor analysis, applied to records for several traits, have found widespread use, both as a dimension reduction technique and to identify common factors.

Analyses of quantitative genetic data have the added complexity, not shared by other fields, that we want to partition the phenotypic variance into its genetic and non-genetic components, using information on covariances between relatives to do so. For 'standard' multi-trait analyses, considering records for multiple characteristics usually measured only once per individual, little attention has been paid to estimation assuming any structural relationship. However, for traits with repeated records, measured potentially infinitely many times along some continuous scale such as time, also referred to as function-valued or functional traits, there has been substantial interest in modelling changing covariances along the associated trajectory, both at the genetic and environmental level. Following suggestions by Kirkpatrick *et al.* [29], covariances were described through covariance functions, based on orthogonal polynomials of the continuous covariable, which were generally estimated to reduced rank, *i.e.* involving polynomials of lower order than the number of points on the trajectory. It was shown that covariance functions could be estimated efficiently fitting mixed linear models which included sets of RR coefficients for all levels of random effects, and estimating the coefficients of the covariance functions as covariances among RR coefficients [39]. A recent review of applications in animal breeding is given by Schaeffer [48].

The parametric structures, widely used elsewhere [21, 57], have been considered as well in genetic analyses of such data. Pletcher and Geyer [46]

teamed parametric correlation functions involving low numbers of parameters with variance functions to accommodate heterogeneous variances, to model genetic and non-genetic covariances between records taken at different ages, dubbing the resulting models 'character process' models. Alternatively, 'hybrid' models have been described, which model genetic trajectories through sets of RR coefficients, but assume a parametric correlation structure, *e.g.* an auto-regressive or an ante-dependence structure, for the within subject covariances [8, 40]. However, applications of the latter two types of model have been limited so far [1, 2, 18].

Eigenvalues and -vectors of covariance matrices *per se* have received relatively little attention. Hayes and Hill [12] used the so-called 'canonical decomposition', which involves simultaneous diagonalisation of two covariance matrices through an eigen-decomposition, to examine sampling properties of genetic selection indexes. Subsequently, Hayes and Hill [13] proposed a modification of eigenvalues on the canonical scale ('bending') to constrain non-positive definite estimates of genetic covariance matrices to the parameter space. In mixed model analyses, both genetic evaluation and variance component estimation, the transformation given by the associated matrix of eigenvectors has been used extensively to reduce computational requirements.

For models without additional random effects, with all traits measured on all animals and with equal design matrices, the canonical transformation reduces the multivariate analysis to a series of univariate analyses; see Jensen and Mao [23] for a review. Extensions to models with multiple random effects and missing data have been considered, in particular for genetic evaluation [6, 7, 31]. Whilst it was recognised that canonical indices, *i.e.* linear combinations of the original variables which are independent of each other and explain maximum amounts of genetic variation, have an interpretation in their own right (see Meyer [36] for an example), this has found little use. For instance, Wiggans and Goddard [55] developed six linear combinations of test day records on dairy cows for milk, fat and protein yields to represent most genetic variation in 60 measures based on a canonical decomposition of the genetic covariance matrix.

Popularity of RR analyses for function-valued traits has sparked interest in the eigendecomposition of covariance matrices of RR coefficients, especially at the genetic level. Eigenfunctions are the 'infinite-dimensional' equivalents to the eigenvectors, and can be estimated by evaluating the eigenvectors of coefficient matrices for the range of covariables fitted [29]. Genetic eigenfunctions then describe the expected deformation of the mean trajectory due to selection, *i.e.* the change in any point along the trajectory likely to occur

when selecting on estimated breeding values for this eigenfunction; see Kirkpatrick *et al*. [29] and Kirkpatrick and Lofsvold [28] for further details. This implies that selection strategies to optimise changes along the complete curve are most easily determined considering the eigenfunctions directly [27, 54]. Furthermore, we are not interested in estimates of the RR coefficients themselves, only in the resulting curve. Hence, as suggested by Kirkpatrick and Meyer [30], it is logical to estimate the values of eigenfunctions directly. This can be done by a simple reparameterisation of the model, as shown in this paper.

Estimating eigenvectors and -values of covariance matrices is equivalent to PC analysis [25]. PC analysis goes back to Hotelling [17] and Pearson [44] (both reprinted in Bryant and Atchley [3]). PC analysis is widely used as a dimension reduction technique, but so far has had only limited applications in quantitative genetic analyses. Eigenvectors define independent linear functions of the variables considered, the so-called principal components, which successively explain the maximum amount of variation, measured by the corresponding eigenvalues. This implies that for a given number of components considered, PCs minimise the mean square of deviations or, conversely, approximate the multivariate data most accurately. It follows that PCs with variances (eigenvalues) close to zero contribute virtually no information to the analysis, that is not already contained in the PCs with large eigenvalues. Hence, these components can be ignored, resulting in, by and large, equivalent analyses involving fewer variables, *i.e.* traits or RR coefficients, and often reduced sampling variation. Reducing the dimension of analyses by considering fewer variables can considerably decrease the associated computational requirements.

This is applicable for both genetic evaluation and variance component estimation, and to all models involving multiple, correlated effects, *i.e.* 'standard' multivariate analyses as well as RR analyses. In terms of variance components, such analyses provide estimates of covariance matrices which are of correspondingly reduced rank and are smoothed. Kirkpatrick and Meyer [30] considered a derivative-free algorithm to estimate reduced rank genetic covariance matrices by restricted maximum likelihood (REML), using a PC parameterisation. James *et al*. [19] applied a PC model at the phenotypic level to sparse, function-valued data, using maximum likelihood estimation.

This paper extends the approach of Kirkpatrick and Meyer [30] to algorithms using derivatives of the likelihood, in particular the so-called 'average information' (AI) algorithm, and gives an application for a multivariate analysis of 'live' carcass traits in beef cattle, recorded by ultrasound scanning.

2. METHOD

2.1. Linear model

In the usual notation, let

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{1}$$

denote the mixed linear model, with $\mathbf{y}$ the vector of observations on N animals, $\boldsymbol{\beta}$ the vector of fixed effects, $\boldsymbol{\alpha}$ the random vector of additive genetic effects for N_A animals (including any parents without records), $\boldsymbol{\gamma}$ the random vector of animals' permanent environmental effects (including any non-additive genetic effects), $\boldsymbol{\varepsilon}$ the vector of residual errors, and $\mathbf{X}, \mathbf{Z}$ and $\mathbf{W}$ the incidence matrices pertaining to $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, respectively.

Assume $\mathbf{y}$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are ordered according to traits within animals, and let $\boldsymbol{\Sigma}_A$, $\boldsymbol{\Sigma}_R$ and $\boldsymbol{\Sigma}_E$ denote the $k \times k$ matrices of additive genetic, permanent environmental and temporary environmental covariances for the k traits. Temporary environmental effects are usually assumed independently distributed, i.e. $\boldsymbol{\Sigma}_E = \mathrm{D}iag\{\sigma_i^2\}$.
Assumed covariances are then

$$Var(\boldsymbol{\alpha}) = \mathbf{A} \otimes \boldsymbol{\Sigma}_A = \mathbf{G} \tag{2}$$

$$Var(\boldsymbol{\gamma}) = \mathbf{I}_N \otimes \boldsymbol{\Sigma}_R = \mathbf{T} \tag{3}$$

$$Var(\boldsymbol{\varepsilon}) = \sum_{j=1}^{N}{}^{+} \boldsymbol{\Sigma}_E^{n_j} = \mathbf{R} \tag{4}$$

with $\mathbf{A}$ the numerator relationship matrix between animals, $\mathbf{I}_N$ an identity matrix of size N, $\boldsymbol{\Sigma}_E^{n_j}$ the submatrix of Σ_E for the jth animal with $n_j \leq k$ observations, obtained by deleting any rows and columns pertaining to traits not recorded, and '$\otimes$' and 'Σ^{+}' denoting the direct matrix product and sum, respectively. Covariances between random effects $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}$ are taken to be zero. This gives

$$Var(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{W}\mathbf{T}\mathbf{W}' + \mathbf{R} = \mathbf{V}. \tag{5}$$

2.1.1. Implicit permanent environmental effects

In the simplest case, (1) represents a 'standard' multivariate model for k traits, with rows of $\mathbf{Z}$ and $\mathbf{W}$ containing a single non-zero element of unity only. If there are single records per trait, temporary environmental effects and

permanent environmental effects due to the individual cannot be separated. This implies that we need to amalgamate $\mathbf{W}\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ in (1) into a single effect $\boldsymbol{\varepsilon}^{\star} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ with variance

$$Var(\boldsymbol{\varepsilon}^{\star}) = \sum_{j=1}^{N}{}^{+} \left(\boldsymbol{\Sigma}_{R}^{n_j} + \boldsymbol{\Sigma}_{E}^{n_j}\right) = \mathbf{R}^{\star} \tag{6}$$

where $\boldsymbol{\Sigma}_{R}^{n_j}$ is the submatrix of $\boldsymbol{\Sigma}_R$ corresponding to $\boldsymbol{\Sigma}_{E}^{n_j}$, and phenotypic variance $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}^{\star}$.

Alternatively, we might choose to fit an equivalent model to (1) which accounts for permanent environmental covariances between repeated records for the same individual through the residual covariances, instead of explicitly fitting a random effect $\boldsymbol{\gamma}$ to model permanent environmental effects. This might be preferable if the number of equations in a mixed model analysis are of concern, if only some of the traits have repeated observations, or if we want to impose a parametric correlation structure on $\boldsymbol{\Sigma}_R$ [40].

2.1.2. Extensions

Extensions of (1) and the algorithm described below to more complicated models including additional random effects, such as genetic or permanent environmental maternal effects, are straightforward. Note, that (1) encompasses RR models, where we describe the trajectories for individuals over the range of continuous covariable(s) through corresponding sets of regression coefficients. For instance, in a univariate RR model, we might fit k_A genetic coefficients α_{ij} for each animal and k_R permanent environmental coefficients γ_{ij} for each individual with records, with covariance matrices $\boldsymbol{\Sigma}_A$ and $\boldsymbol{\Sigma}_R$, respectively. In contrast to the simple multivariate model, incidence matrices for RR models have elements corresponding to the vector of covariates, *i.e.* are considerably less sparse. For example, if we modelled the additive genetic trajectory for each animal through k_A coefficients, regressing on Legendre polynomials of age at recording, each row of $\mathbf{Z}$ would have k_A non-zero elements consisting of Legendre polynomials of degree 0 to $k_A - 1$, evaluated for the age at which the corresponding record was taken. Computations, however, are analogous.

2.2. Reparameterisation

Consider a covariance matrix $\boldsymbol{\Sigma}$, of dimension $k \times k$. An eigenvalue decomposition of $\boldsymbol{\Sigma}$ gives

$$\boldsymbol{\Sigma} = \mathbf{E}\,\boldsymbol{\Lambda}\,\mathbf{E}' \tag{7}$$

with $\mathbf{\Lambda}$ the diagonal matrix of eigenvalues of $\mathbf{\Sigma}$, λ_i for $i = 1,\ldots,k$, and $\mathbf{E} = (\mathbf{e}_1|\mathbf{e}_2|\ldots|\mathbf{e}_k)$ the pertaining matrix of eigenvectors $\mathbf{e}_i$. For a particular eigenvalue λ_i, the corresponding $\mathbf{e}_i$ is only determined to proportionality. Hence, standard procedures for eigenvalue decomposition usually report the $\mathbf{e}_i$ standardised to a length of unity, so that $\mathbf{E}$ is orthonormal. In addition, eigenvalues and -vectors are usually given in descending order of magnitude of the λ_i, and this is assumed to hold in the following.

If $\mathbf{\Sigma}$ represented the covariance matrix of a vector of variables $\mathbf{v}$, the linear function $\mathbf{e}'_1\mathbf{v}$ with variance λ_1 would be the combination of the original variables which explained the maximum amount of variation. Similarly, with $\mathbf{E}$ orthogonal, $\mathbf{E}'\mathbf{v}$ with covariance matrix $\mathbf{\Lambda}$ gives k uncorrelated variables with the ith new variable explaining most variation, given counterparts 1 to $i-1$. Hence, if we considered only the first m eigenvectors, we would obtain m linear combinations $\mathbf{E}'_m\mathbf{v}$ capturing the maximum of the original variation (with $\mathbf{E}_m$ the $k \times m$ submatrix of columns $1,\ldots,m$ of $\mathbf{E}$). This is the principle underlying the use of principal components as a 'dimension reduction' technique. Moreover, if $\lambda_{m+1},\ldots,\lambda_k$ are close to zero, the 'reproduced' matrix

$$\mathbf{\Sigma}^\star = \mathbf{E}_m\mathbf{\Lambda}_m\mathbf{E}'_m = \sum_{i=1}^{m} \lambda_i\mathbf{e}_i\mathbf{e}'_i \qquad (8)$$

with $\mathbf{\Lambda}_m$ the submatrix of $\mathbf{\Lambda}$ corresponding to $\mathbf{E}_m$, is a close approximation of $\mathbf{\Sigma}$ which has rank m and is smoothed.

Assume $\mathbf{\Sigma}$ above stood for $\mathbf{\Sigma}_A$. We can then reparameterise (1) to

$$\mathbf{y} = \mathbf{X}\mathbf{\beta} + \mathbf{Z}(\mathbf{I}_{N_A} \otimes \mathbf{E}_m)(\mathbf{I}_{N_A} \otimes \mathbf{E}'_m)\mathbf{\alpha} + \mathbf{W}\mathbf{\gamma} + \mathbf{\varepsilon} = \mathbf{X}\mathbf{\beta} + \mathbf{Z}^\circ\mathbf{\alpha}^\circ + \mathbf{W}\mathbf{\gamma} + \mathbf{\varepsilon} \quad (9)$$

with $Var(\mathbf{\alpha}^\circ) = \mathbf{A} \otimes \mathbf{\Lambda}_m$. A corresponding reparameterisation, based on the decomposition of $\mathbf{\Sigma}_R$, can be applied to $\mathbf{\gamma}$. For $m = k$, (9) is an equivalent model to (1). Otherwise, *i.e.* for $m < k$, it has reduced dimension, considering the first m principal components only.

In other words, estimating $\mathbf{\alpha}^\circ$ instead of $\mathbf{\alpha}$ in a mixed model analysis directly yields estimates of breeding values for the linear combinations of the original variables given by the first m eigenvectors. This is of particular interest when (1) represents a RR model, so that the eigenvectors of $\mathbf{\Sigma}_A$ describe deformations of the mean trajectory due to selection [29]. Misztal *et al.* [42] considered the parameterisation shown in (9) for RR analyses of test-day records on dairy cattle, but failed to recognise the interpretation of $\mathbf{\alpha}^\circ$.

Alternatively, as suggested by Kirkpatrick and Meyer [30], we can reparameterise (1) so that the random effects fitted have variances of unity.

$$\mathbf{Q}_m = \mathbf{E}_m(\mathbf{\Lambda}_m)^{1/2} = \left(\sqrt{\lambda_1}\mathbf{e}_1|\ldots|\sqrt{\lambda_m}\mathbf{e}_m \right) \text{ so that } \mathbf{\Sigma}^\star = \mathbf{Q}_m\mathbf{Q}'_m \qquad (10)$$

$\mathbf{Q}_m$ is the $k \times m$ matrix of the first $m \leq k$ eigenvectors of the $k \times k$ covariance matrix $\mathbf{\Sigma}$, scaled by the square root of the corresponding eigenvalues. Note that the columns of $\mathbf{Q}_m$ do not have a norm of unity and, for $m < k$, $\mathbf{Q}_m$ is not square. For $m = k$, $\mathbf{Q}^{-1} = \mathbf{\Lambda}^{-1/2}\mathbf{E}' = \mathbf{\Lambda}^{-1}\mathbf{Q}'$.

Introducing subscripts A and R to distinguish between genetic and environmental effects and, for convenience, omitting subscript m in the following, we can then reparameterise (1) to

$$\begin{aligned}
\mathbf{y} &= \mathbf{X\beta} + \mathbf{Z}(\mathbf{I}_{N_A} \otimes \mathbf{Q}_A)\mathbf{\alpha}^\star + \mathbf{W}(\mathbf{I}_N \otimes \mathbf{Q}_R)\mathbf{\gamma}^\star + \mathbf{\varepsilon} \\
&= \mathbf{X\beta} + \mathbf{Z}^\star\mathbf{\alpha}^\star + \mathbf{W}^\star\mathbf{\gamma}^\star + \mathbf{\varepsilon}
\end{aligned} \tag{11}$$

with $\mathbf{\alpha}^\star$ of length $N_A \times m_A$, $\mathbf{\alpha} = (\mathbf{I}_{N_A} \otimes \mathbf{Q}_A)\mathbf{\alpha}^\star$ and $Var(\mathbf{\alpha}^\star) = \mathbf{A} \otimes \mathbf{I}_{m_A}$, and $\mathbf{\gamma}^\star$ of length $N \times m_R$, $\mathbf{\gamma} = (\mathbf{I}_N \otimes \mathbf{Q}_R)\mathbf{\gamma}^\star$ and $Var(\mathbf{\gamma}^\star) = \mathbf{I}_N \otimes \mathbf{I}_{m_R}$. Again, for $m_A = k_A$ and $m_R = k_R$, (11) is equivalent to (1).

2.3. Variance component estimation

Kirkpatrick and Meyer [30] described how estimates of the parameters defining the dispersion structure for model (11) can be estimated by REML, using a derivative-free algorithm. They employed a step-wise procedure to maximise the likelihood with respect to only the elements of the ith eigenvectors in the ith step, to reduce the dimension of search and thus the number of likelihood evaluations required. This was teamed with a final maximisation with respect to all parameters. Using a graphical interpretation, *i.e.* considering changes in the directions and lengths of the eigenvectors through small rotations, allowed a maximisation with respect to the rotation angles, thus keeping the number of parameters to be estimated low.

In the following we describe REML algorithms to estimate the 'variance components' for model (11) utilising information on derivatives of the likelihood. The AI-REML algorithm, described by Johnson and Thompson [24] and Gilmour *et al*. [9] for univariate scenarios, utilises the average of observed and expected information in a Newton-Raphson type iterative estimation scheme. The average information is equivalent to second derivatives of the data part of the likelihood only, and is thus considerably easier to compute than either of the two former values. Multivariate implementations have been described, for instance, by Jensen *et al*. [22] and Meyer [38]. Whilst most implementations relied on computing the inverse of the coefficient matrix in the mixed model equations (MME) [9, 22, 24], the latter utilised a Cholesky factorisation of the coefficient matrix, and this approach is adopted in the following.

2.3.1. *Parameters to be estimated*

In contrast to other parameterisations, the 'variance components' for the random effects in (11) are incorporated into their design matrices, $\mathbf{Z}^{\star}$ and $\mathbf{W}^{\star}$. Let $\boldsymbol{\theta}$ of length p denote the vector of (variance) parameters to be estimated. On the original scale, there were $k(k + 1)/2$ covariance components for each covariance matrix $\boldsymbol{\Sigma}$ of size $k \times k$. On first glance, it might appear that parameterisation (11) might have increased the number of parameters to be estimated, unless m was substantially smaller than k. Whilst each matrix $\mathbf{Q}$ has mk elements, it is a matrix of eigenvectors, albeit scaled, and we require eigenvectors $\mathbf{e}_i$ for $i = 1, \ldots, m$ to be orthogonal. This implies that we have k parameters for the first eigenvector $\mathbf{q}_1 = \sqrt{\lambda_1}\mathbf{e}_1$. For the second eigenvector, $\mathbf{q}_2 = \sqrt{\lambda_2}\mathbf{e}_2$, however, there are only $k - 1$ 'free' elements, with the kth element determined by the restriction that $\mathbf{q}_1'\mathbf{q}_2 = \sqrt{\lambda_1\lambda_2}\mathbf{e}_1'\mathbf{e}_2 = 0$. Analogously, for the ith column of $\mathbf{Q}$, there are only $k - i + 1$ elements to be estimated, with the remaining $i - 1$ elements fully determined by columns $1, \ldots, i - 1$ of $\mathbf{Q}$ which are already estimated. Let elements 1 to $k - i + 1$ of $\mathbf{q}_i$ be the elements to be estimated. Elements $k - i + 2$ to k are then obtained as solutions to the $i - 1$ equations given by $\mathbf{q}_r'\mathbf{q}_i = 0$ for $r = 1, \ldots, i - 1$.

$$\begin{pmatrix} q_{i,k-i+2} \\ q_{i,k-i+3} \\ \vdots \\ q_{i,k} \end{pmatrix} = - \begin{pmatrix} q_{1,k-i+2} & q_{1,k-i+3} & \cdots & q_{1,k} \\ q_{2,k-i+2} & q_{2,k-i+3} & \cdots & q_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{i-1,k-i+2} & q_{i-1,k-i+3} & \cdots & q_{i-1,k} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{r=1}^{k-i+1} q_{1,r}q_{i,r} \\ \sum_{r=1}^{k-i+1} q_{2,r}q_{i,r} \\ \vdots \\ \sum_{r=1}^{k-i+1} q_{k,r}q_{i,r} \end{pmatrix} \quad (12)$$

with $q_{i,r}$ the rth element of $\mathbf{q}_i$ (commas shown here for better separation of subscripts only). Hence, the number of parameters to be estimated for $\mathbf{Q}$ is $m(2k - m + 1)/2 \leq k(k + 1)/2$, *i.e.* for a reduced dimension parameterisation, the number of variables describing the covariance structure is reduced by $(k-m)(k-m+1)/2$. The reduction increases quadratically with the difference in dimension. In specific cases, the matrix to be inverted in (12) can become singular, which implies that the corresponding elements of $\mathbf{q}_i$ are undetermined.

The number of parameters to be estimated in total is then

$$p = m_A (2k_A - m_A + 1)/2 + m_R (2k_R - m_R + 1)/2 + m_E \quad (13)$$

where m_E is the number of distinct elements of $\boldsymbol{\Sigma}_E$. Parameters are the elements of $\mathbf{Q}_A$, $q_{A_{ij}}$ with $i = 1, m_A$ and $j = 1, k_A - m_A + 1$, the elements of $\mathbf{Q}_R$, $q_{R_{ij}}$ with $i = 1, m_R$ and $j = 1, k_R - m_R + 1$, and the temporary environmental

variances, $\sigma_1^2, \ldots, \sigma_{m_E}^2$. Note, that there are no constraints on the parameter space for the former. Variances σ_i^2 have to be positive, but estimating $\log(\sigma_i^2)$ instead of σ_i^2 yields a parameterisation, for which maximising the likelihood is an unconstrained optimisation problem.

For a univariate RR analysis, m_E might be as low as 1 when measurement error variances are assumed homogeneous. For standard multivariate analyses of k traits, usually $m_E = k$ and $k_R \leq k - 1$. If there are only single records per trait, there is no information to separate temporary and permanent environmental covariances, i.e. $m_E = 0$ and $m_R = k$. Alternatively, we might impose a structure of $\mathbf{Q}_R\mathbf{Q}_R' + \mathrm{D}iag\{\sigma_i^2\}$ on $\mathbf{\Sigma}_R + \mathbf{\Sigma}_E$. This facilitates estimation of temporary environmental variances even though there are no repeated records per animal, but m_R must be chosen small enough, so that $m_R(m_R + 1)/2 + m_E \leq k(k + 1)/2$.

2.3.2. Likelihood

The MME for (11) are then

$$
\begin{pmatrix}
\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}^\star & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W}^\star \\
\mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{Z}^\star + \mathbf{A}^{-1} \otimes \mathbf{I}_{m_A} & \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{W}^\star \\
\mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{Z}^\star & \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{W}^\star + \mathbf{I}_N \otimes \mathbf{I}_{m_R}
\end{pmatrix}
\begin{pmatrix}
\hat{\boldsymbol{\beta}} \\
\hat{\boldsymbol{\alpha}}^\star \\
\hat{\boldsymbol{\gamma}}^\star
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\
\mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{y} \\
\mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{y}
\end{pmatrix} \quad (14)
$$

or $\mathbf{C}\hat{\mathbf{u}} = \mathbf{r}$.

Whilst the part of the coefficient matrix $\mathbf{C}$ arising from the covariances among random effects is sparser than in the corresponding MME for (1), it has to be borne in mind that $\mathbf{Z}^\star$ and $\mathbf{W}^\star$ can be considerably denser than matrices $\mathbf{Z}$ and $\mathbf{W}$, although there is a 'trade-off' as the covariance matrices among random effects proportional to identity matrices generate fewer links between random effects levels. Hence computational savings in manipulating the MME arise mainly from the reduction in the number of equations when $m_A < k_A$ or $m_R < k_R$.

The REML log likelihood (log $\mathcal{L}$) can be written in terms of the MME [10]. On the original scale,

$$- 2 \log \mathcal{L} = \text{const.} + N_A \log |\boldsymbol{\Sigma}_A| + k_A \log |\mathbf{A}| + N \log |\boldsymbol{\Sigma}_R|$$
$$+ \log |\mathbf{R}| + \log |\mathbf{C}| + \mathbf{y}'\mathbf{Py}. \quad (15)$$

For the reparameterised model (11),

$$-2 \log \mathcal{L} = \text{const.} + m_A \log |\mathbf{A}| + \log |\mathbf{R}| + \log |\mathbf{C}| + \mathbf{y}'\mathbf{Py} \quad (16)$$

with $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$ a projection matrix, so that $\mathbf{Py} = \mathbf{R}^{-1}\hat{\boldsymbol{\varepsilon}}$ with $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}^\star\hat{\boldsymbol{\alpha}}^\star - \mathbf{W}^\star\hat{\boldsymbol{\gamma}}^\star$. Contributions to log $\mathcal{L}$ from the covariance matrices among random effects, $\boldsymbol{\Sigma}_A$ and $\boldsymbol{\Sigma}_R$, are now contained in $\log |\mathbf{C}|$ and $\mathbf{y}'\mathbf{Py}$.

For a given model, $m_A \log |\mathbf{A}|$ is a constant and is not required when maximising log $\mathcal{L}$. However, $\log |\mathbf{A}|$ is readily obtained as a by-product when setting up $\mathbf{A}^{-1}$. With $\mathbf{R}$ blockdiagonal for animals, its (log) determinant is readily determined considering blocks for one animal at a time. If there are animals with the same combinations of traits (ages) recorded, the number of determinants which need to be evaluated is the number of combinations, and $\log |\mathbf{R}|$ is equal to the sum of products of individual determinants and the respective number of animals [37].

As shown by Graser *et al*. [10], both $\mathbf{y}'\mathbf{Py}$ and $\log |\mathbf{C}|$ can be evaluated simultaneously by factoring the so-called mixed model matrix (MMM) $\mathbf{M}$, which is the coefficient matrix in the MME augmented by the vector of right hand sides and its transpose, and the weighted sum of squares of the data. A Cholesky factorisation requires the matrix to be factored to be positive definite. If multiple, fixed effects are fitted, $\mathbf{X}$ is generally not of full rank, resulting in a matrix $\mathbf{M}$ which is not positive definite. Hence it is assumed in the following, without changing notation, that, if applicable, $\mathbf{X}$ has been replaced by a full rank submatrix (or that the factorisation procedure has been modified to accommodate matrices of reduced rank).

$$\mathbf{M} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}^\star & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W}^\star & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{Z}^\star + \mathbf{A}^{-1} \otimes \mathbf{I}_{m_A} & \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{W}^\star & \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{Z}^\star & \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{W}^\star + \mathbf{I}_N \otimes \mathbf{I}_{m_R} & \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z}^\star & \mathbf{y}'\mathbf{R}^{-1}\mathbf{W}^\star & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}.$$
$$(17)$$

If $\boldsymbol{\gamma}$ is not fitted, the corresponding rows and columns in (14) and (17) are omitted, and $\mathbf{R}$ is replaced by $\mathbf{R}^\star$.

Let $\mathbf{L}$ denote the Cholesky factor of $\mathbf{M}$, *i.e.* $\mathbf{M} = \mathbf{LL}'$, with diagonal elements l_{rr} for $r = 1, \ldots, M$. Terms required in (16) can then be calculated as [41]

$$\log |\mathbf{C}| = \sum_{r=1}^{M-1} \log(l_{rr}) \quad \text{and} \quad \mathbf{y}'\mathbf{Py} = l_{MM}^2. \tag{18}$$

Employing sparse matrix techniques together with an appropriate ordering of rows and columns of $\mathbf{M}$, $\log \mathcal{L}$ can be evaluated for large data sets and models of analyses.

2.3.3. Maximising the likelihood

One of the most widely used methods to optimise non-linear functions is the Newton-Raphson (NR) algorithm. Let $\hat{\boldsymbol{\theta}}^t$ be an estimate of $\boldsymbol{\theta}$ from the tth iterate. An update is then given by

$$\hat{\boldsymbol{\theta}}^{t+1} = \hat{\boldsymbol{\theta}}^t - \kappa^t \left(-\mathbf{I}(\hat{\boldsymbol{\theta}}^t) + \tau^t \mathbf{I}_p \right)^{-1} \mathbf{g}(\hat{\boldsymbol{\theta}}^t) \tag{19}$$

where $\mathbf{I}(\hat{\boldsymbol{\theta}}^t)$ and $\mathbf{g}(\hat{\boldsymbol{\theta}}^t)$ are the matrix of second derivatives of $\log \mathcal{L}$ and the vector of first derivatives of $\log \mathcal{L}$, respectively, both evaluated at $\hat{\boldsymbol{\theta}}^t$, and k^t and τ^t are parameters of the optimisation procedure.

If second derivatives in (19) are replaced by their expected values, the algorithm is referred to as Fisher's Method of Scoring. Similarly, for the AI-REML algorithm, the average of observed and expected values is used instead of the second derivatives in (19). Calculation of derivatives of $\log \mathcal{L}$ with respect to the parameters of the PC parameterisation is described in the Appendix.

For $\tau^t = 0$ and $k^t = 1$, (19) reduces to the standard NR algorithm. Generally, however, it is used with some modifications to avoid 'overshooting' or to enforce an increase in $\log \mathcal{L}$. $k^t \leq 1$ is a step size scaling factor. Jennrich and Sampson [20] suggested a 'step halving' procedure, *i.e.*, for $\tau^t = 0$ to try successive values of k^t of 1, 0.5, 0.25...until a value of $\hat{\boldsymbol{\theta}}^{t+1}$ was found which increased $\log \mathcal{L}$.

For $\tau^t \neq 0$ and $k^t = 1$, (19) gives Marquardt [35]'s modification, resulting in a search step which is intermediate between a NR step, and a step in the method of steepest descent (ascent). If required, a simple search for an optimal value of either τ^t or k^t can be carried out to maximize the increase in $\log \mathcal{L}$ achieved. This is especially advisable when observed second derivatives rather than AI are used, as the ratio of computational efforts required for simple function evaluations and calculation of second derivatives is wide.

Other, more sophisticated, modifications of the NR algorithm, aimed at making it more robust and efficient enough to cope with starting values far from the optimum and nonconvex regions, are described by Nocedahl and Wright [43] (Chap. 6).

3. APPLICATION

The method described is illustrated using records for 'carcass traits' measured by live ultrasound scanning for a herd of Australian beef cattle. Traits recorded were eye muscle area (EMA; in cm^2), fat depth at the 12/13th rib (RIB; in mm) and at the P8 site (P8; in mm), and percentage intra-muscular fat (IMF; in % $\times$ 10). Treating records on heifers and bulls as different traits yielded eight traits in the analysis. Ages at recording from 300 to 700 days were considered. After basic edits, the data comprised 20 171 records on 2335 heifers and 3702 bulls, *i.e.* 5605 animals in total. Generally, all four measures for an animal were taken at the same time. However, IMF recording was introduced some time after the other traits. Hence, only 57% of bulls and 69% of heifers had all four traits recorded, with most of the remainder having records for P8, RIB and EMA. In addition, there was a small number of animals with other combinations of traits, due to missing observations or deletion of dubious records. Table I gives further details.

The model of analysis was a simple animal model, fitting contemporary groups (CG), birth type effects and a dam age class effect (heifers *vs*. cows) as cross-classified fixed effects, and age at recording and age of dam as linear and quadratic covariables. CG were defined as herd-management group-sex-date of recording subclasses, with further subdivisions so that each CG comprised ages at recording no more than 60 days apart. Animals' additive genetic effects were the only random effects considered. Extracting pedigree information for up to four generations backwards yielded 2439 ancestors to be included after 'pruning' of pedigrees, *i.e.* a total of 8044 animals in the analysis. Animals with records were progeny of 239 sires and 2284 dams.

REML estimates of covariance components among the eight traits were obtained fitting 1, 2, ..., 8 principal components for genetic effects, using an AI-REML algorithm to locate the maximum of the likelihood. This was supplemented by simple, derivative-free search steps at convergence or when AI-REML failed to improve log $\mathcal{L}$, although first derivatives different from zero suggested that a maximum had not been attained.

Residual covariances matrices were modelled as full rank matrices throughout. As traits were measured on two distinct subsets of animals, there were

Table I. Characteristics of the data structure (P8: fat depth at P8 site, RIB: fat depth at 12/13th rib, EMA: eye muscle area, and IMF: percentage intramuscular fat).

	Heifers				Bulls			
	P8 (mm)	RIB (mm)	EMA (cm^2)	IMF ($\% \times 10$)	P8 (mm)	RIB (mm)	EMA (cm^2)	IMF ($\% \times 10$)
No. of records	2335	2334	2335	1634	3236	3235	3235	1827
Mean	6.808	5.071	55.723	38.182	3.724	2.864	73.613	16.311
S.D.[a]	3.132	2.780	7.579	22.232	1.918	1.113	12.005	17.096
No. CG[b]	57	57	57	42	117	117	117	66

[a] Standard deviation; [b] contemporary group subclasses.

no environmental covariances between traits 1 to 4 and traits 5 to 8. Hence, four eigenvectors of length 4 were fitted separately for each of the two diagonal blocks of the residual covariance matrix, resulting in only 20 parameters for residuals to be estimated (16 zero covariances). REML forms of Akaike's (AIC) and Schwarz' Bayesian (BIC) information criterion [56] were calculated for each analysis. In addition, 28 corresponding bivariate analyses amongst all pairs of traits were carried out. Results from the latter were pooled using the 'iterative summing of expanded part matrices' procedure of Mäntysaari [34], as implemented by Henshall and Meyer [14].

Table II summarises characteristics of the analyses. In essence, fitting $m < 8$ PCs reduces the mixed model equations and computational demands per likelihood evaluation to those of a corresponding m-variate, full rank analysis. As requirements increase quadratically with the dimension of analysis (m), even a small reduction resulted in a substantial decrease in the number of operations needed for numerical factorisation of the MMM and its derivatives. With the design matrices for random effects for the PC parameterisation comprising the eigenvectors and thus being markedly denser than for standard multivariate analyses, more elements in the MMM arise from the data part. However, this is balanced by substantially fewer elements contributed by the inverses of covariance matrices of random effects. In our example, fitting eight principal components, there were 1 322 063 non-zero off-diagonal elements in the lower triangle of **M** from the data part and 164 568 from the inverse of the genetic covariance matrix. Corresponding figures for the usual parameterisation were 571 675 and 1 541 776. Whilst this resulted in fewer elements in **M** for the PC parameterisation, numbers of non-zero elements in its Cholesky factor and operation counts per factorisation were of comparable magnitude.

Likelihoods increased significantly with the number of PCs fitted up to a fit of 6. Similarly, AIC was lowest for 6 PCs, suggesting this was the best

Table II. Characteristics of analyses fitting $1, 2, \ldots, 8$ principal components and a standard multivariate analysis (MV8), together with maximum log likelihood ($\log \mathcal{L}$) and Akaike (AIC) and Bayesian (BIC) information criteria.

Analysis	Number of							
	rows in MMM[b]	elements	elements in Ch. factor	operations[a] (in 10^6)	para-meters	$\log \mathcal{L}$ +30 400	AIC –61 000	BIC –61 300
Fit 1	8731	182 043	674 816	122.1	28	–197.01	250.03	170.61
Fit 2	16 775	351 590	1 339 004	292.7	35	–112.42	94.84	70.56
Fit 3	24 819	526 751	2 151 401	565.4	41	–80.88	43.77	66.76
Fit 4	32 863	707 517	3 074 970	951.6	46	–61.43	14.87	77.25
Fit 5	40 907	893 888	4 136 889	1469.4	50	–56.41	12.81	106.70
Fit 6	48 951	1 085 864	5 124 419	2243.3	53	–51.82	9.65	127.17
Fit 7	56 995	1 283 445	6 776 224	3056.1	55	–50.86	11.72	145.00
Fit 8	65 039	1 486 631	8 257 550	4090.5	56	–50.85	13.71	154.86
MV 8	65 039	2 086 463	8 220 700	4235.1	56	–50.85	13.71	154.86

[a] For numerical factorisation of the MMM; each operation is equivalent to one multiplication and addition; [b] mixed model matrix.

Table III. Estimates of eigenvalues of the genetic covariance matrix (λ_i) together with their total and the proportion of variance explained by the first three eigenvalues, for analyses fitting $1, 2, \ldots, 8$ principal components and a pooled covariance matrix from bivariate analyses (Bi).

Anal.	Eigenvalues								Total	Prop. var. (%)		
	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	$\sum \lambda$	λ_1	λ_2	λ_3
Fit 1	78.454								78.5	100.0		
Fit 2	83.187	13.428							96.6	86.1	13.9	
Fit 3	114.894	15.359	3.307						133.6	86.0	11.5	2.5
Fit 4	155.786	19.393	3.262	0.744	0.744				179.2	86.9	10.8	1.8
Fit 5	157.275	19.691	3.533	0.830	0.330				181.7	86.6	10.8	1.9
Fit 6	157.740	19.244	14.123	1.737	1.135	0.187			194.2	81.2	9.9	7.3
Fit 7	157.740	19.252	14.123	1.733	1.136	0.184	0.066		194.2	81.2	9.9	7.3
Fit 8	158.144	19.202	14.065	1.734	1.133	0.182	0.066	0.000	194.5	81.3	9.9	7.2
Bi	155.882	18.123	14.374	1.693	0.743	0.145	0.036	0.000	191.0	81.6	9.5	7.5

fitting model. Involving a more stringent penalty for the number of parameters estimated, a more parsimonious model fitting the first 3 PCs only would be chosen on the basis of BIC. Whilst estimates of the fourth (λ_4) and fifth (λ_5) eigenvalue, shown in Table III, were low for analyses fitting 4 and 5 PCs, respectively, there appeared to be some repartitioning of variation, resulting in estimates of λ_4 and λ_5 being higher for analyses fitting 6 or more PCs. Estimates of eigenvalues from pooled bivariate analyses agreed well with those from analyses fitting 6 or more PCs.
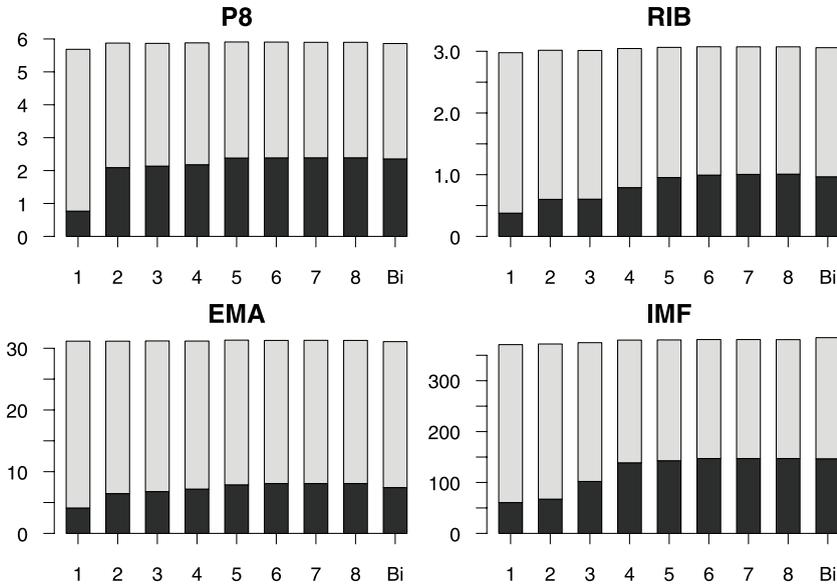
Figure 1. Estimates of genetic (■) and residual (▨) variance components for traits measured on heifers (P8: fat depth at P8 site, RIB: fat depth at 12/13th rib, EMA: eye muscle area, and IMF: percentage intra-muscular fat), for analysis fitting $1, 2, \ldots, 8$ principal components together with pooled estimates from bivariate analyses (Bi).

As the sum of genetic eigenvalues shows, the total variance attributed to genetic effects increased with the number of PCs fitted, implying a biased partitioning, *i.e.* that residual covariances have been overestimated, for the lowest orders of fit. This is further illustrated in Figure 1, which shows that estimates of phenotypic variances for individual traits were essentially the same for all analyses, while estimates of genetic variances increased and estimates of the residual variances decreased with increasing number of PCs considered (up to 6). Figure 2 shows corresponding changes in estimates of selected genetic correlations. Results suggest that 6 PCs may be required to estimate covariances between the eight traits adequately, but, conversely, that the first three of these PCs suffice to explain 98.4% of genetic variation between animals.

4. DISCUSSION

Multi-dimensional analyses involving more than a few correlated genetic effects are inherently difficult. Typically, computational requirements are large. In addition, sampling variances and probabilities of parameter estimates outside the boundaries of the parameter space, increase with the dimension of
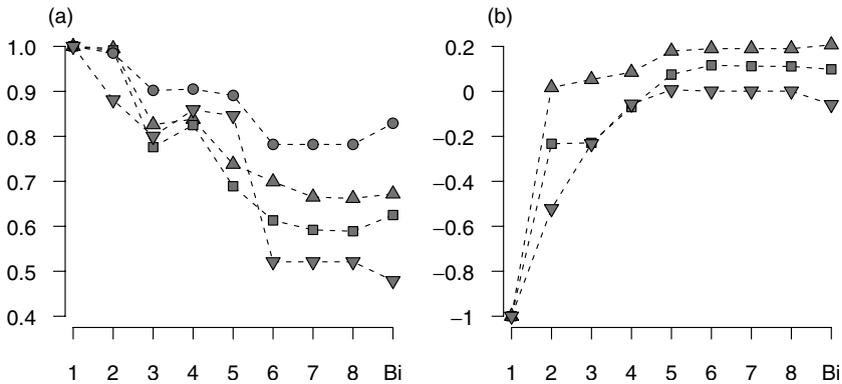
Figure 2. Estimates of genetic correlations between the same trait measured on different sexes (a) and between eye muscle area and the other traits measured on heifers (b) for analysis fitting $1, 2, \ldots, 8$ principal components together with pooled estimates from bivariate analyses (Bi).
(▲: fat depth at P8 site, ■: fat depth at 12/13th rib, ▼: percentage intra-muscular fat, and ●: eye muscle area).

analysis and magnitude of genetic correlations [15]. The model proposed is capable of alleviating both problems. Estimating only the most important principal components allows a more parsimonious parameterisation than standard multivariate analyses, while still capturing the bulk of variation and smoothing estimates of the corresponding covariance matrices. In a small simulation study for a RR model, Kirkpatrick and Meyer [30] showed that the first eigenvector and -value tended to be estimated accurately regardless of the number of principal components fitted, and that this dominated the sampling errors in the associated covariance function. Further work is required to evaluate the sampling behaviour in more practical situations.

In addition, computational demands can be decreased markedly, when fitting reduced numbers of PCs. This could prove especially advantageous for RR analyses, where regression on higher order polynomials has frequently resulted in 'end-of-range' problems. Lower order, segmented polynomials, *i.e.* splines, may provide a more flexible and less sensitive alternative. In particular, Rice and Wu [47] recommended the so-called 'B-splines' [4] as base functions, and Torres and Quaas [53] presented an application for test day records of dairy cows. A drawback of spline models is that, depending on the number of segments or 'knots' used, the number of regression coefficients and thus parameters to be estimated is often increased compared to polynomial models. However, when the first few PCs suffice to estimate covariance functions, reduced rank RR analyses using spline base functions do not need to be any more

complex than current, full rank analyses using polynomials. James *et al.* [19] considered this scenario at a phenotypic level.

The model and algorithm presented are closely related to the factor-analytic (FA) model and AI-REML algorithms considered by Smith *et al.* [49] and Thompson *et al.* [51]. In particular, if specific variances are assumed zero, the FA model collapses into a reduced rank model [51]. However, both authors assumed the random effect with FA structure to be distributed proportionally to an identity matrix, and necessary constraints on the number of parameters to be estimated were imposed by setting the $k(k-1)/2$ elements in the upper triangle of the matrix of factor loadings to zero [49]. In addition, both algorithms relied on calculation of $\mathbf{C}^{-1}$, the inverse of the coefficient matrix in the MME.

4.1. Convergence behaviour

Initial experience with the algorithm described, using an AI-REML algorithm together with simple 'step halving', showed that a modification of the step size was required for most iterations, sometimes substantially so. This lead to slower convergence than the quadratic convergence expected and often observed; see Hofer [16] for a summary of literature results. For the example presented, estimating all eight principal components required more than twice as many iterates than the corresponding standard, multi-trait analysis. Hence, the PC parameterisation is most advantageous for a reduced rank fit, where any increase in iterations required is off-set by a smaller number of parameters to be estimated, and much lower computational requirements to evaluate $\log \mathcal{L}$ and its derivatives.

Comparing various transformations of covariance matrices, Pinheiro and Bates [45] encountered slower convergence in obtaining maximum likelihood estimates for a parameterisation using eigenvalues and Givens rotation angles of the matrix than for the alternatives considered. The NR algorithm and its modifications are based on a quadratic expansion of the likelihood function about its maximum, *i.e.* they work best when the shape of the likelihood surface is unimodal and close to quadratic. Little is known about the shape of the likelihood surface for the principal component parameterisation suggested. James *et al.* [19] commented briefly on difficulties of maximising $\log \mathcal{L}$ due to non-convexity of the optimisation problem, but gave no further details.

4.2. Variance matrix non-linear in vector of parameters

Results given by Hofer [16] by and large referred to models in which the covariance matrix of observations ($\mathbf{V}$) was linear in the parameters to be

estimated. In this case, the AI is an exact average of observed and expected values of second derivatives of $\log \mathcal{L}$ [9]. Otherwise, contributions from non-zero second derivatives of $\mathbf{V}$ are omitted. No reports of convergence rates for $\mathbf{V}$ non-linear in $\boldsymbol{\theta}$ were available.

For a full rank fit, derivatives of $\log \mathcal{L}$ with respect to the elements of $\mathbf{Q}$ can be obtained by evaluating derivatives with respect to covariance components σ_{rs} (elements of $\boldsymbol{\Sigma}$) and then transforming these to the eigenvector scale. Using the chain rule of differentiation for parameters $\theta_v = q_{rs}$ and $\theta_w = q_{tu}$ and covariances $\phi_x = \sigma_{r's'}$ and $\phi_y = \sigma_{t'u'}$ [33]

$$\frac{\partial^2 \log \mathcal{L}}{\partial \theta_v \partial \theta_w} = \sum_{x=1}^{k(k+1)/2} \sum_{y=1}^{k(k+1)/2} \frac{\partial \phi_x}{\partial \theta_v} \frac{\partial \phi_y}{\partial \theta_w} \frac{\partial^2 \log \mathcal{L}}{\partial \phi_x \partial \phi_y} + \sum_{x=1}^{k(k+1)/2} \frac{\partial^2 \phi_x}{\partial \theta_v \partial \theta_w} \frac{\partial \log \mathcal{L}}{\partial \phi_x}. \quad (20)$$

The first component in the right hand side of (20) is the average information given by (A.13). The second part arises from the fact that some second derivatives of α^{xy} with respect to the elements of the corresponding eigenvectors, q_{rs}, are non-zero. As (20) clearly shows, this component is the more important the larger first derivatives are, *i.e.* the further we are away from the maximum of the likelihood. With $\sigma_{xy} = \sum_{r=1}^{m} q_{rx} q_{ry}$, partial derivatives in (20) are simple to evaluate, non-zero values being $\partial \sigma_s^2 / \partial q_{rs} = 2q_{rs}$ and $\partial \sigma_{sy} / \partial q_{rs} = q_{ry}$, and $\partial^2 \sigma_s^2 / (\partial q_{rs})^2 = 2$ and $\partial^2 \sigma_{sy} / \partial q_{rs} \partial q_{ry} = 1$, for $r, s, y = 1, \ldots, k$.

4.3. Alternative parameterisation

In constructing our algorithm to maximise the (log) likelihood, we have adopted the conventional approach of taking derivatives of $\log \mathcal{L}$ with respect to the parameters to be estimated. Alternatives exist and may be advantageous.

Kirkpatrick and Meyer [30] considered a search strategy to exploit the orthogonality of PCs: after first estimating the last (mth) PC conditional on those $(1, \ldots, m-1)$ already estimated, they suggested a final optimisation step consisting of small rotations and perturbations of all PCs simultaneously. For m PCs, this involved maximisation of $\log \mathcal{L}$ for $m(m-1)/2$ rotation angles and m perturbation factors, *i.e.* $m(m+1)/2$ parameters. Earlier, Juga and Thompson [26] teamed rotations and a canonical transformation in a derivative-free search scheme for bivariate analyses.

It is straightforward to rewrite the mixed model and MMM, in terms of fixed eigenvectors and variable rotation angles and perturbation factors as the parameters to be estimated. Derivatives of $\log \mathcal{L}$ with respect to the new variables are readily evaluated in the same manner as described above for the elements of

eigenvectors, and the likelihood can be maximised accordingly. Further research is planned to implement such algorithm and examine its efficacy.

For consistency, the algorithm has been presented assuming a reparameterisation to scaled eigenvectors for all covariance components to be estimated, other than temporary environmental variances. This is not mandatory, and in some instances we may want to combine the eigenvector parameterisation for genetic effects with a different form for other random effects. For covariance matrices fitted to full rank, it may be advantageous to compute the average information for the covariance components themselves first, and then transform these to derivatives with respect to the elements of the scaled eigenvectors (shown above), as this would capture the non-linear part of (20) ignored otherwise.

Indeed, for random effects where we are not interested in the eigen-structure *per se*, another parameterisation may be preferable. In particular, for the Cholesky decomposition, $\mathbf{\Sigma} = \mathbf{L}_{\Sigma}\mathbf{L}_{\Sigma}'$, a parameterisation to estimate the elements of the lower triangular matrix $\mathbf{L}_{\Sigma}$, taking log values of the diagonals to remove constraints on the parameter space, has been used [33, 41, 45] and been found to yield good convergence rates [11, 45, 52]. Pivoting on the largest diagonal, this parameterisation can be used to estimate reduced rank covariance matrices, by estimating the first m columns of $\mathbf{L}_{\Sigma}$ only [39]. Again, more work is required to determine the best strategy and optimisation procedure to be used.

5. CONCLUSIONS

Estimation of genetic principal components directly is appealing. It has several advantages over the estimation of individual genetic effects and their covariance components, including reduction in the number of parameters to be considered, sampling variation and computational requirements. Apart from offering greater parsimony, it provides readily interpretable results characterising the patterns of genetic covariation in multiple dimension. It is expected to facilitate routine analyses considering more than a few traits simultaneously.

REFERENCES

[1] Albuquerque L.G., Meyer K., Estimates of genetic covariance functions for growth assuming a parametric correlation structure for permanent environmental effects, CDROM Seventh World Congr. Genet. Appl. Livest. Prod., 2002, Communication No. 2–21.

[2] Bermejo J.L., Roehe R., Rave G., Kalm E., Comparison of linear and nonlinear functions and covariance structures to estimate feed intake pattern in growing pigs, Livest. Prod. Sci. 82 (2003) 15–26.

[3] Bryant E.H., Atchley W.R., Multivariate Statistical Methods: Within-Groups Covariation, Vol. 2 of Benchmark Papers in Systematic and Evolutionary Biology, Dowden, Hutchinson & Ross, Inc.; Halsted Press, Stroudsburg, PA, 1975.

[4] de Boor C., A Practical Guide to Splines, Vol. 27 of Springer Series in Applied Mathematics, Springer-Verlag, New York, 2nd edn., 2001.

[5] Diggle P.J., Liang K.Y., Zeger S.L., Analysis of Longitudinal Data, Oxford Science Publications, Clarendon Press, Oxford, 1994.

[6] Ducrocq V., Besbes B., Solution of multiple trait animal models with missing data on some traits, J. Anim. Breed. Genet. 110 (1993) 81–92.

[7] Ducrocq V., Chapuis H., Generalising the use of the canonical transformation for the solution of multivariate mixed model equations, Genet. Sel. Evol. 29 (1997) 205–224.

[8] Foulley J.L., Jaffrézic F., Robert-Granié C., EM-REML estimation of covariance parameters in Gaussian mixed models for longitudinal data analysis, Genet. Sel. Evol. 32 (2000) 129–141.

[9] Gilmour A.R., Thompson R., Cullis B.R., Average Information REML, an efficient algorithm for variance parameter estimation in linear mixed models, Biometrics 51 (1995) 1440–1450.

[10] Graser H.U., Smith S.P., Tier B., A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood, J. Anim. Sci. 64 (1987) 1362–1370.

[11] Groeneveld E., A reparameterisation to improve numerical optimisation in multivariate REML (co)variance component estimation, Genet. Sel. Evol. 26 (1994) 537–545.

[12] Hayes J.F., Hill W.G., A reparameterisation of a genetic index to locate its sampling properties, Biometrics 36 (1980) 237–248.

[13] Hayes J.F., Hill W.G., Modifications of estimates of parameters in the construction of genetic selection indices ('bending'), Biometrics 37 (1981) 483–493.

[14] Henshall J.M., Meyer K., "PDMATRIX" – Programs to make matrices positive definite, CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod., 2002, Communication No. 28–12.

[15] Hill W.G., Thompson R., Probabilities of non-positive definite between-group or genetic covariance matrices, Biometrics (1978) 429–439.

[16] Hofer A., Variance component estimation in animal breeding: a review, J. Anim. Breed. Genet. 115 (1998) 247–265.

[17] Hotelling H., Analysis of a complex of statistical variables into principal components, J. Educ. Psych. 24 (1933) 417–441.

[18] Jaffrézic F., White I.M.S., Thompson R., Visscher P.M., Contrasting models for lactation curve analysis, J. Dairy Sci. 84 (2002) 968–975.

[19] James G.M., Hastie T.J., Sugar C.A., Principal component models for sparse functional data, Biometrika 87 (2000) 587–602.

[20] Jennrich R.I., Sampson P.F., Newton-Raphson and related algorithms for maximum likelihood variance component estimation, Technometrics 18 (1976) 11–17.

[21] Jennrich R.I., Schluchter M.D., Unbalanced repeated-measures models with structured covariance matrices, Biometrics 42 (1986) 805–820.

[22] Jensen J., Mäntysaari E.A., Madsen P., Thompson R., Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information, J. Ind. Soc. Agr. Stat. 49 (1997) 215–225.

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

[24] Johnson D.L., Thompson R., Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information, J. Dairy Sci. 78 (1995) 449–456.

[25] Joliffe I.T., Principal Component Analysis, Springer Series in Statistics, Springer-Verlag, New York, 1986.

[26] Juga J., Thompson R., A derivative-free algorithm to estimate bivariate (co)variance components using canonical transformations and estimated rotations, Acta Agric. Scand. A 42 (1992) 191–197.

[27] Kirkpatrick M., Selection on age dependent traits, CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod., 2002, Communication No. 19–06.

[28] Kirkpatrick M., Lofsvold D., Measuring selection and constraint in the evolution of growth, Evolution 46 (1992) 954–971.

[29] Kirkpatrick M., Lofsvold D., Bulmer M., Analysis of the inheritance, selection and evolution of growth trajectories, Genetics 124 (1990) 979–993.

[30] Kirkpatrick M., Meyer K., Simplified analysis of complex phenotypes: Direct estimation of genetic principal components, Genetics 168 (2004) 2295–2306.

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

[32] Lindsey J.K., Models for Repeated Measurements, Vol. 19, Oxford Statistical Science Series, Oxford University Press, 2nd edn., 1999.

[33] Lindstrom M.J., Bates D.M., Newton-Raphson and EM algorithms for linear mixed effects models for repeated-measures data, J. Amer. Stat. Ass. 83 (1988) 1014–1022.

[34] Mäntysaari E.A., Derivation of multiple trait reduced random regression (RR) model for the first lactation test day records of milk, protein and fat, in: 50th Annual Meeting, Europ. Ass. Anim. Prod., (1999) mimeo., 8p.

[35] Marquardt D.W., An algorithm for least squares estimation of nonlinear parameters, SIAM J. 11 (1963) 431–411.

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

[37] Meyer K., Estimating variances and covariances for multivariate animal models by restricted maximum likelihood, Genet. Sel. Evol. 23 (1991) 67–83.

[38] Meyer K., An "average information" restricted maximum likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices, Genet. Sel. Evol. 29 (1997) 97–116.

[39] Meyer K., Estimating covariance functions for longitudinal data using a random regression model, Genet. Sel. Evol. 30 (1998) 221–240.

[40] Meyer K., Estimating genetic covariance functions assuming a parametric correlation structure for environmental effects, Genet. Sel. Evol. 33 (2001) 557–585.

[41] Meyer K., Smith S.P., Restricted maximum likelihood estimation for animal models using derivatives of the likelihood, Genet. Sel. Evol. 28 (1996) 23–49.

[42] Misztal I., Strabel T., Jamrozik J., Mäntysaari E.A., Meuwissen T.H.E., Strategies for estimating the parameters needed for different test-day models, J. Dairy Sci. 83 (2000) 1125–1134.

[43] Nocedahl J., Wright S.J., Numerical Optimization, Springer Series in Operations Research, Springer-Verlag, New York, Berlin Heidelberg, 1999.

[44] Pearson K., On lines and planes of closest fit to systems of points in space, Phil. Mag. 2 (1901) 559–572.

[45] Pinheiro J.C., Bates D.M., Unconstrained parameterizations for variance-covariance matrices, Stat. Comp. 6 (1996) 289–296.

[46] Pletcher S.D., Geyer C.J., The genetic analysis of age-dependent traits: Modeling the character process, Genetics 153 (1999) 825–835.

[47] Rice J.A., Wu C.O., Nonparametric mixed effects models for unequally sampled noisy curves, Biometrics 57 (2001) 253–259.

[48] Schaeffer L.R., Application of random regression models in animal breeding, Livest. Prod. Sci. 86 (2004) 35–45.

[49] Smith A.B., Cullis B.R., Thompson R., Analysing variety by environment data using multiplicative mixed models and adjustments for spatial field trends, Biometrics 57 (2001) 1138–1147.

[50] Smith S.P., Differentiation of the Cholesky algorithm, J. Comp. Graph. Stat. 4 (1995) 134–147.

[51] Thompson R., Cullis B.R., Smith A.B., Gilmour A.R., A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models, Austr. New Zeal. J. Stat. 45 (2003) 445–459.

[52] Thompson R., Mäntysaari E.A., Prospects for statistical methods in dairy cattle breeding, Interbull Bull 20 (1999) 70–77.

[53] Torres R.A.A., Quaas R.L., Determination of covariance functions for lactation traits on dairy cattle using random-coefficient regressions on B-splines, J. Anim. Sci. 79 (2001) Suppl.1, 112, abstract.

[54] van der Werf J.H.J., Optimising selection for traits along trajectories, CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod., 2002, Communication No. 16–07.

[55] Wiggans G.R., Goddard M.E., A computationally feasible test day model for genetic evaluation of yield traits in the United States, J. Dairy Sci. 80 (1997) 1795–1800.

[56] Wolfinger R.D., Covariance structure selection in general mixed models, Comm. Stat. - Simul. Comp. 22 (1993) 1079–1106.

[57] Wolfinger R.D., Heterogeneous variance-covariance structures for repeated measures data, J. Agric. Biol. Env. Stat. 1 (1996) 205–230.

APPENDIX

A.1. First derivatives

First derivatives of the log likelihood are

$$-2\frac{\partial \log \mathcal{L}}{\partial \theta_t} = \frac{\partial \log |\mathbf{R}|}{\partial \theta_t} + \frac{\partial \log |\mathbf{C}|}{\partial \theta_t} + \frac{\partial \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial \theta_t} \tag{A.1}$$

$$= tr\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_t}\right) + tr\left(\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_t}\right) + \mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_t}\mathbf{P}\mathbf{y} \text{ for } t = 1,\ldots,p.$$

As outlined by Meyer and Smith [41], the last two terms required in (A.1) can be evaluated simultaneously, using the 'automatic differentiation' of a Cholesky factor proposed by Smith [50] (see also Nocedahl and Wright [43], Chap. 7). This requires the corresponding derivatives of $\mathbf{M}$ to be specified.

Let $\partial \mathbf{M}/\partial \theta_t = (\partial \mathbf{L}/\partial \theta_t)(\partial \mathbf{L}'/\partial \theta_t)$ with $\partial l_{rr}/\partial \theta_t$ the diagonal elements of the derivative of the Cholesky factor, $\partial \mathbf{L}/\partial \theta_t$. Derivatives of $\log |\mathbf{C}|$ and $\mathbf{y}'\mathbf{P}\mathbf{y}$ are then [41]

$$\frac{\partial \log |\mathbf{C}|}{\partial \theta_t} = 2\sum_{r=1}^{M-1} l_{rr}^{-1}\frac{\partial l_{rr}}{\partial \theta_t} \quad \text{and} \quad \frac{\partial \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial \theta_t} = 2l_{MM}\frac{\partial l_{MM}}{\partial \theta_t}. \tag{A.2}$$

A.1.1. Derivatives of design matrices and $\log |\mathbf{R}|$

For a covariance matrix $\mathbf{\Sigma} = \mathbf{Q}\mathbf{Q}'$, first derivatives with respect to the elements of $\mathbf{Q}$ are

$$\frac{\partial \mathbf{\Sigma}}{\partial q_{rs}} = \frac{\partial \mathbf{Q}}{\partial q_{rs}}\mathbf{Q}' + \mathbf{Q}\frac{\partial \mathbf{Q}'}{\partial q_{rs}} = D^{rs}\mathbf{Q}' + \mathbf{Q}D^{rs\prime} = \mathbf{H}^{rs} \tag{A.3}$$

with $\mathbf{D}^{rs}$ a matrix of the same size as $\mathbf{Q}$, $k \times m$, with a single non-zero element of unity in position rs. Hence $\mathbf{H}^{rs}$ is a symmetric matrix of the same size as $\mathbf{\Sigma}$, $k \times k$, with elements $h_{ss} = 2q_{rs}$ and $h_{st} = h_{ts} = q_{rt}$ for $t \neq s = 1, \ldots, k$, and zero otherwise.

Introducing subscripts to distinguish between matrices $\mathbf{D}^{rs}$ pertaining to $\mathbf{Q}_A$ and $\mathbf{Q}_R$, the non-zero derivatives of the design matrices are

$$\frac{\partial \mathbf{Z}^{\star}}{\partial q_{Ars}} = \frac{\partial \mathbf{Z}\,(\mathbf{I}_{N_A} \otimes \mathbf{Q}_A)}{\partial q_{Ars}} = \mathbf{Z}\left(\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{rs}\right) \tag{A.4}$$

$$\text{for} \quad r = 1, \ldots, m_A; s = 1, \ldots, k_A - r + 1;$$

$$\frac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}} = \frac{\partial \mathbf{W}\,(\mathbf{I}_N \otimes \mathbf{Q}_R)}{\partial q_{Rrs}} = \mathbf{W}\left(\mathbf{I}_N \otimes \mathbf{D}_R^{rs}\right) \tag{A.5}$$

$$\text{for} \quad r = 1, \ldots, m_R; s = 1, \ldots, k_R - r + 1.$$

As for $\log|\mathbf{R}|$, its derivatives can be evaluated for one animal or one combination of records at a time.

$$tr\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_t}\right) = \sum_{j=1}^{N} tr\left((\mathbf{\Sigma}_E^{n_j})^{-1}\frac{\partial \mathbf{\Sigma}_E^{n_j}}{\partial \log\left(\sigma_s^2\right)}\right) = \sum_{j=1}^{N} n_j^s \tag{A.6}$$

where n_j^s is the number of records for the jth animal which have temporary environmental variance σ_s^2.

A.1.2. Derivatives of M

With the design matrices $\mathbf{Z}^{\star}$ and $\mathbf{W}^{\star}$ incorporating the 'variance components' for the random effects fitted, derivatives of $\mathbf{M}$ for the reparameterised model (11) have a less simple structure than those for model (1) [41]. For elements of $\mathbf{Q}_A$ and $\mathbf{Q}_R$, the derivatives are

$$\frac{\partial \mathbf{M}}{\partial q_{Ars}} =$$

$$\begin{pmatrix}
0 & \mathbf{X}'\mathbf{R}^{-1}\dfrac{\partial \mathbf{Z}^{\star}}{\partial q_{Ars}} & 0 & 0 \\[2ex]
\dfrac{\partial \mathbf{Z}^{\star\prime}}{\partial q_{Ars}}\mathbf{R}^{-1}\mathbf{X} & \dfrac{\partial \mathbf{Z}^{\star\prime}}{\partial q_{Ars}}\mathbf{R}^{-1}\mathbf{Z}^{\star} + \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\dfrac{\partial \mathbf{Z}^{\star}}{\partial q_{Ars}} & \dfrac{\partial \mathbf{Z}^{\star\prime}}{\partial q_{Ars}}\mathbf{R}^{-1}\mathbf{W}^{\star} & \dfrac{\partial \mathbf{Z}^{\star\prime}}{\partial q_{Ars}}\mathbf{R}^{-1}\mathbf{y} \\[2ex]
0 & \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\dfrac{\partial \mathbf{Z}^{\star}}{\partial q_{Ars}} & 0 & 0 \\[2ex]
0 & \mathbf{y}'\mathbf{R}^{-1}\dfrac{\partial \mathbf{Z}^{\star}}{\partial q_{Ars}} & 0 & 0
\end{pmatrix}$$

$$\tag{A.7}$$

and

$$\frac{\partial \mathbf{M}}{\partial q_{Rrs}} =$$

$$\begin{pmatrix} 0 & 0 & \mathbf{X}'\mathbf{R}^{-1}\dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}} & 0 \\[2ex] 0 & 0 & \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}} & 0 \\[2ex] \dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}}\mathbf{R}^{-1}\mathbf{X} & \dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}}\mathbf{R}^{-1}\mathbf{Z}^{\star} & \dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}}\mathbf{R}^{-1}\mathbf{W}^{\star} + \mathbf{W}^{\star\prime}\mathbf{R}^{-1}\dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}} & \dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}}\mathbf{R}^{-1}\mathbf{y} \\[2ex] 0 & 0 & \mathbf{y}'\mathbf{R}^{-1}\dfrac{\partial \mathbf{W}^{\star}}{\partial q_{Rrs}} & 0 \end{pmatrix}.$$

$$\tag{A.8}$$

$\partial \mathbf{M}/\partial \log(\sigma_s^2)$, the corresponding matrix of derivatives for the temporary environmental variances, has the same form as $\mathbf{M}$ given in (17) but with each $\mathbf{R}^{-1}$ replaced by its derivative $\partial \mathbf{R}^{-1}/\partial \log(\sigma_s^2)$, for $s = 1, \ldots, m_E$.

$$\frac{\partial \mathbf{R}^{-1}}{\partial \log\left(\sigma_s^2\right)} = -\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \log\left(\sigma_s^2\right)}\mathbf{R}^{-1} \tag{A.9}$$

$$= \sum_{j=1}^{N}{}^{+} - (\mathbf{\Sigma}_E^{n_j})^{-1}\frac{\partial \mathbf{\Sigma}_E^{n_j}}{\partial \log\left(\sigma_s^2\right)}(\mathbf{\Sigma}_E^{n_j})^{-1}.$$

For individual j with n_j records, the submatrix of (A.9) is $-\sigma_s^{-2}\mathbf{D}_j^s$, where $\mathbf{D}_j^s$ is a matrix of size $n_j \times n_j$, which has a diagonal element of unity for each record which has temporary environmental variance σ_s^2, and elements of zero otherwise.

A.1.3. Implicit permanent environmental effects

Again, if $\boldsymbol{\gamma}$ is not fitted, (A.7) and $\partial \mathbf{M}/\partial \log(\sigma_s^2)$ are reduced accordingly, omitting rows and columns pertaining to $\boldsymbol{\gamma}$, and $\mathbf{R}$ is replaced by $\mathbf{R}^{\star}$. In that case, blocks of $\mathbf{R}^{\star}$ are submatrices of $\mathbf{\Sigma}_R + \mathbf{\Sigma}_E$ which is assumed to have structure $\mathbf{Q}_R\mathbf{Q}_R' + \mathrm{D}iag\{\sigma_i^2\}$. Hence, derivatives $\partial \mathbf{R}^{\star}/\partial_{q_{Rrs}}$ are non-zero, with the block for the jth animal equal to

$$-\left(\mathbf{\Sigma}_R^{n_j} + \mathbf{\Sigma}_E^{n_j}\right)^{-1}\left(\mathbf{H}_R^{rs}\right)^{n_j}\left(\mathbf{\Sigma}_R^{n_j} + \mathbf{\Sigma}_E^{n_j}\right)^{-1} \tag{A.10}$$

where $(\mathbf{H}_R^{rs})^{n_j}$ is the submatrix of $\mathbf{H}_R^{rs} = \partial \mathbf{\Sigma}_R/\partial q_{Rrs}$, as defined above (see (A.3)), corresponding to the combination of traits recorded for individual j. Alternatively, we may have assumed $\mathbf{\Sigma}_R + \mathbf{\Sigma}_E = \mathbf{Q}_\mathbf{R}\mathbf{Q}_\mathbf{R}'$ with

$m_R = k_R$. In that case, we do not estimate separate components σ_s^2, *i.e.* $p = (m_A(2k_A - m_A + 1) + k_R(k_R + 1))/2$ and $\partial \mathbf{M}/\partial \sigma_s^2$ do not exist. Similarly, derivatives of $\log |\mathbf{R}^\star|$ are

$$tr\left((\mathbf{R}^\star)^{-1}\frac{\partial \mathbf{R}^\star}{\partial \theta_t}\right) = \sigma_s^2 \sum_{j=1}^{N} tr\left(\left(\mathbf{\Sigma}_R^{n_j} + \mathbf{\Sigma}_E^{n_j}\right)^{-1}\mathbf{D}_j^s\right) \quad \text{for} \quad \theta_t = \log(\sigma_s^2); \quad \text{(A.11)}$$

$$tr\left((\mathbf{R}^\star)^{-1}\frac{\partial \mathbf{R}^\star}{\partial \theta_t}\right) = \sum_{j=1}^{N} tr\left(\left(\mathbf{\Sigma}_R^{n_j} + \mathbf{\Sigma}_E^{n_j}\right)^{-1}(\mathbf{H}_R^{rs})^{n_j}\right) \quad \text{for} \quad \theta_t = q_{R\,rs}. \quad \text{(A.12)}$$

A.2. Average information

As shown by Gilmour *et al.* [9] and Johnson and Thompson [24], the 'average information' is proportional to the second derivatives of the data part of $\log \mathcal{L}$, $\mathbf{y}'\mathbf{Py}$.

$$\frac{\partial^2 \mathbf{y}'\mathbf{Py}}{\partial \theta_t \partial \theta_u} = \mathbf{y}'\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_t}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \theta_u}\mathbf{Py} = \mathbf{b}_t'\mathbf{Pb}_u \quad \text{with} \quad \mathbf{b}_t = \frac{\partial \mathbf{V}}{\partial \theta_t}\mathbf{Py}. \quad \text{(A.13)}$$

A.2.1. Work vectors $\mathbf{b}_t$

For derivatives with respect to elements of the eigenvectors, vectors $\mathbf{b}_t$ can be rewritten in terms of estimates of the corresponding random effects

$$\mathbf{b}_t = \mathbf{Z}\left(\mathbf{I}_{N_A} \otimes \mathbf{H}_A^{rs}(\mathbf{\Sigma}_A^\star)^-\right)\hat{\boldsymbol{\alpha}} = \mathbf{Z}\left(\mathbf{I}_{N_A} \otimes \mathbf{H}_A^{rs}(\mathbf{\Sigma}_A^\star)^-\mathbf{Q}_A\right)\hat{\boldsymbol{\alpha}}^\star$$
$$\text{for } \theta_t = q_{A\,rs}; \quad \text{(A.14)}$$

$$\mathbf{b}_t = \mathbf{W}\left(\mathbf{I}_N \otimes \mathbf{H}_R^{rs}(\mathbf{\Sigma}_R^\star)^-\right)\hat{\boldsymbol{\gamma}} = \mathbf{W}\left(\mathbf{I}_N \otimes \mathbf{H}_R^{rs}(\mathbf{\Sigma}_R^\star)^-\mathbf{Q}_R\right)\hat{\boldsymbol{\gamma}}^\star$$
$$\text{for } \theta_t = q_{R\,rs}. \quad \text{(A.15)}$$

For derivatives with respect to temporary environmental variances

$$\mathbf{b}_t = \left(\sum_{j=1}^{N}{}^+\mathbf{D}^s\right)\hat{\boldsymbol{\varepsilon}} \quad \text{for } \theta_t = \log(\sigma_s^2) \quad \text{(A.16)}$$

i.e., $\mathbf{b}_t$ has elements equal to the estimated residuals for records which are assumed to have error variance σ_s^2 and zero otherwise.

A.2.2. Calculation of average information

As outlined above, $\mathbf{y}'\mathbf{P}\mathbf{y}$ is readily calculated from $\mathbf{y}'\mathbf{R}^{-1}\mathbf{y}$ by factoring the mixed model matrix $\mathbf{M}$. As described in more detail in Meyer [38,40], this can be extended to determine products $\mathbf{b}'_t\mathbf{P}\mathbf{b}_u$.

Define a matrix $\mathbf{B}$ with columns $\mathbf{b}_r$ for $r = 1, \ldots, p$. Replacing each vector $\mathbf{y}$ and $\mathbf{y}'$ in $\mathbf{M}$ with $\mathbf{B}$ and $\mathbf{B}'$, respectively, then yields a matrix $\mathbf{M}_B$ with $M - 1 + p$ rows and columns. Factoring $\mathbf{M}_B$ or, equivalently, 'absorbing' the first $M - 1$ rows and columns into the last p rows and columns of $\mathbf{M}_B$, then replaces $\mathbf{B}'\mathbf{R}^{-1}\mathbf{B}$ with $\mathbf{B}'\mathbf{P}\mathbf{B}$, which, to proportionality, is the average information matrix. As the Cholesky factorisation of the first $M - 1$ rows and columns of $\mathbf{M}_B$ has already been determined in calculating $\log \mathcal{L}$, this is computationally undemanding.

A.2.3. Implicit permanent environmental effects

For $\boldsymbol{\gamma}$ not fitted, derivatives of $\mathbf{R}^\star$ with respect to q_{Rrs} are non-zero, and the equivalent to (A.15) is

$$\mathbf{b}_t = \left(\sum_{j=1}^{N}{}^{+} (\mathbf{H}_R^{rs})^{n_j} \left(\boldsymbol{\Sigma}_R^{n_j} + \boldsymbol{\Sigma}_E^{n_j} \right)^{-1} \right) \hat{\boldsymbol{\varepsilon}} \qquad \text{for } \theta_t = q_{Rrs}. \qquad (A.17)$$

A.3. Second derivatives

Analogously, second derivatives of the log likelihood are

$$-2\frac{\partial^2 \log \mathcal{L}}{\partial\theta_t \partial\theta_u} = \frac{\partial^2 \log |\mathbf{R}|}{\partial\theta_t \partial\theta_u} + \frac{\partial^2 \log |\mathbf{C}|}{\partial\theta_t \partial\theta_u} + \frac{\partial^2 \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial\theta_t \partial\theta_u} \quad \text{for } t, u = 1, \ldots, p. \quad (A.18)$$

As above, the last two terms required in (A.18) can be evaluated simultaneously, taking second derivatives of the a Cholesky factor of $\mathbf{M}$. Again, these can be obtained through 'automatic differentiation' [50], provided the second derivatives of $\mathbf{M}$ can be specified. For $\partial^2\mathbf{M}/(\partial\theta_t \partial\theta_u) = (\partial^2\mathbf{L}/(\partial\theta_t \partial\theta_u))(\partial\mathbf{L}'/(\partial\theta_t \theta_u))$ with diagonal elements $\partial^2 l_{rr}/(\partial\theta_t \partial\theta_u)$ [41],

$$\frac{\partial^2 \log |\mathbf{C}|}{\partial\theta_t \partial\theta_u} = 2 \sum_{r=1}^{M-1} l_{rr}^{-1} \frac{\partial l_{rr}}{\partial\theta_t \partial\theta_u} - l_{rr}^{-2} \frac{\partial l_{rr}}{\partial\theta_t} \frac{\partial l_{rr}}{\partial\theta_u} \qquad \text{and} \qquad (A.19)$$

$$\frac{\partial \mathbf{y}'\mathbf{P}\mathbf{y}}{\partial\theta_t \partial\theta_u} = 2 \left(l_{MM} \frac{\partial^2 l_{MM}}{\partial\theta_t \partial\theta_u} + \frac{\partial l_{MM}}{\partial\theta_t} \frac{\partial l_{MM}}{\partial\theta_u} \right). \qquad (A.20)$$

A.3.1. Derivatives of $\log |\mathbf{R}|$

The first term in (A.18),

$$\frac{\partial^2 \log |\mathbf{R}|}{\partial\theta_t\partial\theta_u} = tr\left(\mathbf{R}^{-1}\frac{\partial^2 \mathbf{R}}{\partial\theta_t\partial\theta_u}\right) - tr\left(\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_t}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_u}\right), \qquad \text{(A.21)}$$

can again be evaluated considering submatrices for one animal at a time. As for the corresponding first derivatives (see (A.6–A.12) above), simplifications of (A.21) exist and are straightforward, but are not developed here.

A.3.2. Derivatives of M

Second derivatives of $\mathbf{M}$ have non-zero elements only in blocks for random effects. In the 4×4 grid representing $\mathbf{M}$, diagonal blocks corresponding to $\boldsymbol{\alpha}^\star$, *i.e.*, in position (2,2), are $\partial^2\mathbf{M}/(\partial q_{A\,rs}\partial q_{A\,tu})$ which expand to

$$\left(\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{rs'}\right)\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{tu}\right) + \left(\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{tu'}\right)\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \qquad \text{(A.22)}$$
$$\left(\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{rs}\right).$$

Similarly, in position (3,3), diagonal non-zero diagonal blocks $\partial^2\mathbf{M}/(\partial q_{R\,rs}\partial q_{R\,tu})$ are

$$\left(\mathbf{I}_N \otimes \mathbf{D}_R^{rs'}\right)\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}\left(\mathbf{I}_N \otimes \mathbf{D}_R^{tu}\right) + \left(\mathbf{I}_N \otimes \mathbf{D}_R^{tu'}\right)\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} \qquad \text{(A.23)}$$
$$\left(\mathbf{I}_N \otimes \mathbf{D}_R^{rs}\right).$$

In addition, there are non-zero off-diagonal blocks corresponding to $\boldsymbol{\alpha}^\star$ and $\boldsymbol{\gamma}^\star$. Blocks in position (3,2) are

$$\left(\mathbf{I}_N \otimes \mathbf{D}_R^{tu'}\right)\mathbf{W}'\mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{rs}\right) \qquad \text{in} \quad \partial^2\mathbf{M}/(\partial q_{A\,rs}\partial q_{R\,tu}) \quad \text{(A.24)}$$

$$\left(\mathbf{I}_N \otimes \mathbf{D}_R^{rs'}\right)\mathbf{W}'\mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{I}_{N_A} \otimes \mathbf{D}_A^{tu}\right) \qquad \text{in} \quad \partial^2\mathbf{M}/(\partial q_{R\,rs}\partial q_{A\,tu}) \quad \text{(A.25)}$$

and blocks (2,3) are their respective transposes.

Second derivatives of $\mathbf{M}$ with respect to parameters modelling residual co-variances again have the same structure as $\mathbf{M}$, but with $\mathbf{R}^{-1}$ in (17) replaced by [41]

$$\frac{\partial^2\mathbf{R}^{-1}}{\partial\theta_t\partial\theta_u} = \mathbf{R}^{-1}\left(\frac{\partial\mathbf{R}}{\partial\theta_t}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_u} + \frac{\partial\mathbf{R}}{\partial\theta_u}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\theta_t} - \frac{\partial^2\mathbf{R}}{\partial\theta_t\partial\theta_u}\right)\mathbf{R}^{-1}, \qquad \text{(A.26)}$$

with obvious simplifications if $\mathbf{R}$ is a diagonal matrix of temporary environmental variances.

A.3.3. *Implicit permanent environmental effects*

If $\boldsymbol{\gamma}$ is not fitted, second derivatives with respect to the elements of $\mathbf{Q}_R$ can be evaluated using $\mathbf{H}_R^{rs} = \partial \boldsymbol{\Sigma}_R / \partial q_{R\,rs}$ as defined above, and that second derivatives of $\boldsymbol{\Sigma}_R$ with respect to the elements of its scaled eigenvectors are

$$\frac{\partial^2 \boldsymbol{\Sigma}_R}{\partial q_{R\,rs} \partial q_{R\,tu}} = \mathbf{D}_R^{rs} \mathbf{D}_R^{tu\,\prime} + \mathbf{D}_R^{tu} \mathbf{D}_R^{rs\,\prime} = \tilde{\mathbf{H}}_R^{rstu} \tag{A.27}$$

where $\tilde{\mathbf{H}}^{rstu}$ is a symmetric matrix of same size as $\boldsymbol{\Sigma}_R$ with elements $h_{ss} = 2$ for $r = t$ and $s = u$, $h_{su} = h_{us} = 1$ for $r = t$ and $s \neq u$, and zero otherwise.