
Australian & New Zealand Journal of Statistics

Aust. N. Z. J. Stat. **59**(4), 2017, 433–448 doi: 10.1111/anzs.12208

A new REML (parameter expanded) EM algorithm for linear mixed models

S. M. Diffey[1] **A. B. Smith**[1,*] **A. H. Welsh**[2] **and B. R. Cullis**[1]

University of Wollongong, Australian National University

Summary

Linear mixed models are regularly applied to animal and plant breeding data to evaluate genetic potential. Residual maximum likelihood (REML) is the preferred method for estimating variance parameters associated with this type of model. Typically an iterative algorithm is required for the estimation of variance parameters. Two algorithms which can be used for this purpose are the expectation-maximisation (EM) algorithm and the parameter expanded EM (PX-EM) algorithm. Both, particularly the EM algorithm, can be slow to converge when compared to a Newton-Raphson type scheme such as the average information (AI) algorithm. The EM and PX-EM algorithms require specification of the complete data, including the incomplete and missing data. We consider a new incomplete data specification based on a conditional derivation of REML. We illustrate the use of the resulting new algorithm through two examples: a sire model for lamb weight data and a balanced incomplete block soybean variety trial. In the cases where the AI algorithm failed, a REML PX-EM based on the new incomplete data specification converged in 28% to 30% fewer iterations than the alternative REML PX-EM specification. For the soybean example a REML EM algorithm using the new specification converged in fewer iterations than the current standard specification of a REML PX-EM algorithm. The new specification integrates linear mixed models, Henderson's mixed model equations, REML and the REML EM algorithm into a cohesive framework.

Key words: variance components; mixed model equations

1. Introduction

Linear mixed models are routinely applied to biological data. The preferred method for estimating the parameters associated with these models is residual maximum likelihood (REML) (Patterson & Thompson 1971). Many statistical software packages available for the REML estimation of parameters associated with linear mixed models implement a Newton-Raphson type algorithm such as the Fisher scoring algorithm or the AI algorithm (Gilmour, Thompson & Cullis 1995). There are two problems with these types of algorithms. Firstly, successive iterations of these algorithms are not guaranteed to increase the residual log-likelihood function. Secondly, parameter updates may not remain in their parameter space. Either problem may result in the algorithm failing to converge to a solution unless some modification of a Newton-Raphson type algorithm is implemented (see Smyth 1998 for details on one such modification).

*Author to whom correspondence should be addressed.
[1]Centre for Bioinformatics and Biometrics, University of Wollongong, Wollongong, NSW 2522, Australia
[2]Mathematical Sciences Institute, Australian National University, Canberra, ACT 0200, Australia
e-mail: simon@apexbiometry.com

Acknowledgements. Simon Diffey gratefully acknowledges the financial support provided by the Grains Research Development Corporation and the New South Wales Department of Primary Industries while undertaking a PhD programme.

The EM algorithm (Dempster, Laird & Rubin 1977) and the parameter expanded EM (PX-EM) algorithm (Liu, Rubin & Wu 1998) have the desirable properties of monotonic convergence and the parameter updates remaining in the parameter space. Dempster, Laird & Rubin (1977) and Liu, Rubin & Wu (1998) only consider maximum likelihood estimation although REML estimation is briefly mentioned by Dempster, Laird & Rubin (1977) in their variance components example. A REML EM algorithm to estimate the parameters associated with a linear mixed model applied to longitudinal data was considered by Laird & Ware (1982) and Foulley, Jaffrezic & Robert-Cranie (2000). A REML PX-EM algorithm for linear mixed models was considered by Foulley & Van Dyk (2000). In all the aforementioned papers REML estimates of variance parameters were obtained by considering the vector of fixed effects to have a variance tending to infinity. This is the approach briefly described by Dempster, Laird & Rubin (1977).

The purpose of this paper is to present an alternative REML (PX-)EM algorithm for linear mixed models. The derivation of this algorithm can be used as a foundation for key concepts associated with the estimation of linear mixed model variance parameters. These concepts include Henderson's mixed model equations, REML, and iterative schemes. Two examples are considered and in both cases the alternative algorithm converges in fewer iterations then the standard algorithm.

This paper proceeds in four parts. First, we provide statistical background on the linear mixed model, Henderson's mixed model equations (Henderson et al. 1959), and a conditional derivation of REML (Verbyla 1990). Second, we consider two alternative complete data specifications for a REML EM algorithm. One is the complete data specification used in current specifications of a REML EM algorithm for linear mixed models and the other is a REML EM algorithm based on a new incomplete data specification. We outline a proof that this new algorithm will converge to a (local) maximum of the residual log-likelihood function. Third, we consider a REML PX-EM algorithm for linear mixed models using the new specification. Fourth, we illustrate the use of the current and new REML EM and REML PX-EM algorithms for linear mixed models by considering two examples.

2. Statistical background

2.1. Linear mixed model formulation

The linear mixed model can be written as

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{Zu}+\boldsymbol{e}, \tag{1}$$

where $\boldsymbol{y}$ is an $n\times 1$ vector of observed data, $\boldsymbol{\beta}$ is a $t\times 1$ vector of fixed effects and $\boldsymbol{X}$ is its associated design matrix which is assumed to be of full column rank. The $n\times b$ design matrix $\boldsymbol{Z}$ is associated with the vector of random effects $\boldsymbol{u}\sim N\{\boldsymbol{0},\boldsymbol{G}(\boldsymbol{\gamma})\}$. The matrix $\boldsymbol{G}(\boldsymbol{\gamma})$ is positive definite and depends on the parameter vector $\boldsymbol{\gamma}$. It is assumed that $\boldsymbol{e}\sim \mathrm{N}(\boldsymbol{0},\boldsymbol{R})$, where $\boldsymbol{R}=\sigma_e^2\boldsymbol{\Sigma}(\boldsymbol{\phi})$, and the matrix $\boldsymbol{\Sigma}(\boldsymbol{\phi})$ is assumed to be positive definite and depends on the parameter vector $\boldsymbol{\phi}$. We will assume that n is larger than b but recognise that this is not always the case. The joint distribution of $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{e}$ is

$$\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{u}\\ \boldsymbol{e}\end{pmatrix}\sim \mathrm{N}\left\{\begin{pmatrix}\boldsymbol{X\beta}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{pmatrix},\begin{pmatrix}\boldsymbol{H} & \boldsymbol{ZG} & \boldsymbol{R}\\ \boldsymbol{GZ}^\top & \boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{R} & \boldsymbol{0} & \boldsymbol{R}\end{pmatrix}\right\}, \tag{2}$$

where $\boldsymbol{H}=\boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^\top+\boldsymbol{R}$. The superscript $\top$ denotes matrix transpose. The aim is to estimate the parameter vector $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\boldsymbol{\kappa}^\top)^\top$, where $\boldsymbol{\kappa}=(\boldsymbol{\gamma}^\top,\sigma_e^2,\boldsymbol{\phi}^\top)^\top$ is a vector of variance parameters.

2.2. Henderson's mixed model equations

Components of Henderson's mixed model equations are widely used in schemes for computing REML estimates of variance parameters. A discussion on their convenience for this purpose can be found in section 7.6 of Searle, Casella & McCulloch (2006) and references therein.

Henderson's mixed model equations are derived by considering the log joint density function of $\boldsymbol{y}$ and $\boldsymbol{u}$, i.e.

$$\log\{f(\boldsymbol{y},\boldsymbol{u};\boldsymbol{\theta})\}=\log\{f(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\theta})\}+\log\{f(\boldsymbol{u};\boldsymbol{\gamma})\}.$$

Obtaining the conditional distribution $f(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\theta})$ requires a well known result in multivariate normal statistics and the joint distribution given in (2). The conditional distribution $f(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\theta})$ is

$$\boldsymbol{y}|\boldsymbol{u}\sim\mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}u,\boldsymbol{R}).$$

The log joint density function of $\boldsymbol{y}$ and $\boldsymbol{u}$ is therefore (excluding constants)

$$\begin{aligned}\log\{f(\boldsymbol{y},\boldsymbol{u};\boldsymbol{\theta})\}=-\frac{1}{2}\Big[&\log\{\det(\boldsymbol{R})\}+\log\{\det(\boldsymbol{G})\}+\boldsymbol{u}^\top\boldsymbol{G}^{-1}\boldsymbol{u}\\&+(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}u)^\top\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}u)\Big].\end{aligned}\tag{3}$$

Differentiating (3) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{u}$ and equating to zero gives Henderson's mixed model equations

$$\begin{aligned}\boldsymbol{X}^\top\boldsymbol{R}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}^\top\boldsymbol{R}^{-1}\boldsymbol{Z}\tilde{\boldsymbol{u}}&=\boldsymbol{X}^\top\boldsymbol{R}^{-1}\boldsymbol{y},\\\boldsymbol{Z}^\top\boldsymbol{R}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}}+(\boldsymbol{Z}^\top\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}^{-1})\tilde{\boldsymbol{u}}&=\boldsymbol{Z}^\top\boldsymbol{R}^{-1}\boldsymbol{y},\end{aligned}$$

where $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{u}}$ is the best linear unbiased predictor (BLUP) of $\boldsymbol{u}$. When expressed in matrix notation the coefficient matrix of Henderson's mixed model equations can be written as

$$\boldsymbol{C}=\begin{pmatrix}\boldsymbol{X}^\top\boldsymbol{R}^{-1}\boldsymbol{X}&\boldsymbol{X}^\top\boldsymbol{R}^{-1}\boldsymbol{Z}\\\boldsymbol{Z}^\top\boldsymbol{R}^{-1}\boldsymbol{X}&\boldsymbol{Z}^\top\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}^{-1}\end{pmatrix}=\begin{pmatrix}\boldsymbol{C}_{XX}&\boldsymbol{C}_{XZ}\\\boldsymbol{C}_{ZX}&\boldsymbol{C}_{ZZ}\end{pmatrix}$$

with the inverse of the coefficient matrix as

$$\begin{aligned}\boldsymbol{C}^{-1}&=\begin{pmatrix}(\boldsymbol{X}^\top\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}&-(\boldsymbol{X}^\top\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}\\-\boldsymbol{G}^\top\boldsymbol{Z}^\top\boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}&(\boldsymbol{Z}^\top\boldsymbol{S}\boldsymbol{Z}+\boldsymbol{G}^{-1})^{-1}\end{pmatrix}\\&=\begin{pmatrix}\boldsymbol{C}^{XX}&\boldsymbol{C}^{XZ}\\\boldsymbol{C}^{ZX}&\boldsymbol{C}^{ZZ}\end{pmatrix},\end{aligned}$$

where $\boldsymbol{S}=\boldsymbol{R}^{-1}-\boldsymbol{R}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{R}^{-1}$. The matrix $\boldsymbol{C}^{ZZ}$ is often referred to as the prediction error variance. We will define the conditional variances required for computation of the E-step in the REML EM and REML PX-EM algorithm for linear mixed models in terms related to the coefficient matrix of Henderson'smixed model equations.

2.3. A conditional derivation of REML

The preferred method for estimating the variance parameters associated with a linear mixed model is REML. The original reference for REML is the paper by Patterson & Thompson (1971); however the new implementation of a REML EM and REML PX-EM algorithm for linear mixed models is based on the conditional derivation of REML presented by Verbyla (1990). The conditional derivation of REML begins by considering the transformation

$$\boldsymbol{L}^\top \boldsymbol{y} = \begin{pmatrix} \boldsymbol{L}_1^\top \boldsymbol{y} \\ \boldsymbol{L}_2^\top \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix},$$

where $\boldsymbol{L} = (\boldsymbol{L}_1, \boldsymbol{L}_2)$ is a non-singular matrix, $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are $n \times t$ and $n \times (n-t)$ matrices respectively, both of full-column rank and chosen to satisfy $\boldsymbol{L}_1^\top \boldsymbol{X} = \boldsymbol{I}_t$ and $\boldsymbol{L}_2^\top \boldsymbol{X} = \boldsymbol{0}$. The distribution of the transformed data is

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{L}_1^\top \boldsymbol{H} \boldsymbol{L}_1 & \boldsymbol{L}_1^\top \boldsymbol{H} \boldsymbol{L}_2 \\ \boldsymbol{L}_2^\top \boldsymbol{H} \boldsymbol{L}_1 & \boldsymbol{L}_2^\top \boldsymbol{H} \boldsymbol{L}_2 \end{pmatrix} \right\}.$$

The joint distribution of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ can be expressed as the product of the marginal distribution of $\boldsymbol{y}_2$ and the conditional distribution $\boldsymbol{y}_1 | \boldsymbol{y}_2$. Therefore, the log-likelihood function of $\boldsymbol{L}^\top \boldsymbol{y}$ can be expressed as

$$\ell(\boldsymbol{\theta}; \boldsymbol{L}^\top \boldsymbol{y}) = \ell(\boldsymbol{\kappa}; \boldsymbol{y}_2) + \ell(\boldsymbol{\theta}; \boldsymbol{y}_1 | \boldsymbol{y}_2). \tag{4}$$

The vector $\boldsymbol{y}_1$ is of length t and the conditional log-likelihood function $\ell(\boldsymbol{\theta}; \boldsymbol{y}_1 | \boldsymbol{y}_2)$ is used to estimate the fixed effects, after which there is no information left for the estimation of variance parameters. The log-likelihood function associated with the marginal density of $\boldsymbol{\kappa}$, i.e., $\ell(\boldsymbol{\kappa}; \boldsymbol{y}_2)$ is used for the REML estimation of variance parameters and can be expressed as

$$\ell(\boldsymbol{\kappa}; \boldsymbol{y}_2) = -\frac{1}{2}\left[\log\{\det(\boldsymbol{L}_2^\top \boldsymbol{H} \boldsymbol{L}_2)\} + \boldsymbol{y}_2^\top (\boldsymbol{L}_2^\top \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{y}_2 \right] \tag{5}$$

Using the identities (see Verbyla 1990) we can show that

$$\boldsymbol{P} = \boldsymbol{L}_2 (\boldsymbol{L}_2^\top \boldsymbol{H} \boldsymbol{L}_2)^{-1} \boldsymbol{L}_2^\top = \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{H}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{H}^{-1},$$

$$\log\{\det(\boldsymbol{L}_2^\top \boldsymbol{H} \boldsymbol{L}_2)\} = \log\{\det(\boldsymbol{L}^\top \boldsymbol{L})\} + \log\{\det(\boldsymbol{H})\} + \log\{\det(\boldsymbol{X}^\top \boldsymbol{H}^{-1} \boldsymbol{X})\},$$

and then if we ignore the constant $\log\{\det(\boldsymbol{L}^\top \boldsymbol{L})\}$, (5) can be written as

$$\ell(\boldsymbol{\kappa}; \boldsymbol{y}_2) = -\frac{1}{2}\left[\log\{\det(\boldsymbol{H})\} + \log\{\det(\boldsymbol{X}^\top \boldsymbol{H}^{-1} \boldsymbol{X})\} + \boldsymbol{y}^\top \boldsymbol{P} \boldsymbol{y} \right], \tag{6}$$

which is the the form of the residual log-likelihood function most commonly presented.

3. Two alternative incomplete data specifications

3.1. Current derivations of a REML EM algorithm for the linear mixed model

Current derivations of REML EM and REML PX-EM algorithms for linear mixed models are implemented by considering the vector of fixed effects $\boldsymbol{\beta}$ to be a random effect with variance tending to infinity. We will refer to this implementation as the random effects approach. Using this approach we will assume that $\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{B})$ and consider the complete data to be defined as $\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}^\top, \boldsymbol{\beta}^\top, \boldsymbol{u}^\top)^\top$ where the superscript P in parentheses on $\boldsymbol{y}_c$ is used to distinguish this complete data specification from the new complete data specification

to be presented shortly. The random effects approach is the approach briefly described by Dempster, Laird & Rubin (1977) and followed by others including Laird & Ware (1982); Foulley, Jaffrezic & Robert-Cranie (2000) and Foulley & Van Dyk (2000). We define the joint distribution of $\boldsymbol{y}$, $\boldsymbol{\beta}$, $\boldsymbol{u}$, and $\boldsymbol{e}$ as

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\beta} \\ \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{F} & \boldsymbol{XB} & \boldsymbol{ZG} & \boldsymbol{R} \\ \boldsymbol{BX}^\top & \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{GZ}^\top & \boldsymbol{0} & \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{R} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R} \end{pmatrix} \right\}, \tag{7}$$

where

$$\boldsymbol{F} = \boldsymbol{XBX}^\top + \boldsymbol{ZGZ}^\top + \boldsymbol{R} = \boldsymbol{XBX}^\top + \boldsymbol{H}$$

and $\boldsymbol{F}^{-1} = \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{B}^{-1} + \boldsymbol{X}^\top\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{H}^{-1}$. The random effect approach assumes that the variance of $\boldsymbol{\beta}$ tends to infinity, i.e., $\boldsymbol{B}^{-1} \to \boldsymbol{0}$. Therefore

$$\boldsymbol{F}^{-1} \to \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{H}^{-1} = \boldsymbol{P}.$$

We will assume that if $\boldsymbol{B}^{-1} \to \boldsymbol{0}$ then $\boldsymbol{F}^{-1}$ can be approximated by $\boldsymbol{P}$. The complete data log density function can be written as

$$\log\{f(\boldsymbol{y}_c^{(P)}; \boldsymbol{\theta})\} = -\frac{1}{2}\Big[\log\{\det(\boldsymbol{R})\} + \log\{\det(\boldsymbol{G})\} + \boldsymbol{u}^\top\boldsymbol{G}^{-1}\boldsymbol{u} + \boldsymbol{e}^\top\boldsymbol{R}^{-1}\boldsymbol{e}\Big], \tag{8}$$

which is equivalent to the log joint density function of $\boldsymbol{y}$ and $\boldsymbol{u}$ used to derive Henderson's mixed model equations and presented in (3) except we have made the substitution $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu}$.

The E-step of a REML EM algorithm using $\boldsymbol{y}_c^{(P)}$ involves taking the conditional expectation of the complete data log density function in (8) with respect to $\boldsymbol{y}$ and at the current iterate of $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}^{(w)}$, i.e.,

$$\begin{aligned} Q^{(P)}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(w)}) &= E\big[\log\{f(\boldsymbol{y}_c^{(P)}; \boldsymbol{\theta})\} | \boldsymbol{y}; \boldsymbol{\theta}^{(w)}\big] \\ &= E\big[\log\{f(\boldsymbol{e}; \boldsymbol{\theta})\} | \boldsymbol{y}; \boldsymbol{\theta}^{(w)}\big] + E\big[\log\{f(\boldsymbol{u}; \boldsymbol{\gamma})\} | \boldsymbol{y}; \boldsymbol{\theta}^{(w)}\big] \\ &= Q^{(P)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\theta}^{(w)}) + Q^{(P)}(\boldsymbol{\gamma}; \boldsymbol{\theta}^{(w)}). \end{aligned}$$

Computation of $Q^{(P)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\theta}^{(w)})$ and $Q^{(P)}(\boldsymbol{\gamma}; \boldsymbol{\theta}^{(w)})$ requires the conditional distributions $\boldsymbol{e}|\boldsymbol{y}$ and $\boldsymbol{u}|\boldsymbol{y}$ respectively. From (7) it can be shown that

$$\begin{aligned} \boldsymbol{e}|\boldsymbol{y} &\sim \mathrm{N}\left(\boldsymbol{RPy}, \boldsymbol{WC}^{-1}\boldsymbol{W}^\top\right), \\ \boldsymbol{u}|\boldsymbol{y} &\sim \mathrm{N}\left(\boldsymbol{GZ}^\top\boldsymbol{Py}, \boldsymbol{C}^{\boldsymbol{ZZ}}\right), \end{aligned}$$

where $\boldsymbol{W} = (\boldsymbol{X}, \boldsymbol{Z})$ is an $n \times (t + b)$ matrix. Using these two results we can write

$$\begin{aligned} Q^{(P)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\theta}^{(w)}) &= -\frac{1}{2}\big[\log\{\det(\boldsymbol{R})\} + \tilde{\boldsymbol{e}}^{(w)^\top}\boldsymbol{R}^{-1}\tilde{\boldsymbol{e}}^{(w)} + tr\big(\boldsymbol{R}^{-1}\boldsymbol{WC}^{-1^{(w)}}\boldsymbol{W}^\top\big)\big], \\ Q^{(P)}(\boldsymbol{\gamma}; \boldsymbol{\theta}^{(w)}) &= -\frac{1}{2}\big[\log\{\det(\boldsymbol{G})\} + \tilde{\boldsymbol{u}}^{(w)^\top}\boldsymbol{G}^{-1}\tilde{\boldsymbol{u}}^{(w)} + tr\big(\boldsymbol{G}^{-1}\boldsymbol{C}^{\boldsymbol{ZZ}^{(w)}}\big)\big], \end{aligned}$$

where $\tilde{\boldsymbol{e}}^{(w)} = \boldsymbol{R}^{(w)}\boldsymbol{P}^{(w)}\boldsymbol{y}$ and $\tilde{\boldsymbol{u}}^{(w)} = \boldsymbol{G}^{(w)}\boldsymbol{Z}^\top\boldsymbol{P}^{(w)}\boldsymbol{y}$.

The M-step of a REML EM algorithm for linear mixed models using $\boldsymbol{y}_c^{(P)}$ involves maximising $Q^{(P)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\theta}^{(w)})$ and $Q^{(P)}(\boldsymbol{\gamma}; \boldsymbol{\theta}^{(w)})$ with respect to σ_e^2, $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ respectively. Noting that $\boldsymbol{R} = \sigma_e^2\boldsymbol{\Sigma}$ the updating equation for σ_e^2 is

$$\sigma_e^{2^{(w+1)}} = \frac{1}{n}\left[\tilde{\boldsymbol{e}}^{(w)^\top}\boldsymbol{\Sigma}^{-1^{(w)}}\tilde{\boldsymbol{e}}^{(w)} + tr\left(\boldsymbol{\Sigma}^{-1^{(w)}}\boldsymbol{W}\boldsymbol{C}^{-1^{(w)}}\boldsymbol{W}^\top\right)\right].$$

The updating equations for $\phi_{ij}, i=1,\ldots,n, j=1,\ldots,n, i \geqslant j$ and $\gamma_{gh}, g=1,\ldots b, h=1,\ldots,b, g \geqslant h$, depend on the specific form of $\boldsymbol{\Sigma}$ and $\boldsymbol{G}$ but can be found by equating the following partial derivatives to zero,

$$\begin{aligned}
\frac{\partial Q^{(P)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\theta}^{(w)})}{\partial \phi_{ij}} &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_{ij}}\right) - \frac{1}{\sigma_e^2}\tilde{\boldsymbol{e}}^{(w)^\top}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_{ij}}\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{e}}^{(w)}\right. \\
&\quad \left. - \frac{1}{\sigma_e^2}tr\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \phi_{ij}}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\boldsymbol{C}^{-1^{(w)}}\boldsymbol{W}^\top\right)\right], \\
\frac{\partial Q^{(P)}(\boldsymbol{\gamma}; \boldsymbol{\theta}^{(w)})}{\partial \gamma_{gh}} &= -\frac{1}{2}\left[tr\left(\boldsymbol{G}^{-1}\frac{\partial \boldsymbol{G}}{\partial \gamma_{gh}}\right) - \tilde{\boldsymbol{u}}^{(w)^\top}\boldsymbol{G}^{-1}\frac{\partial \boldsymbol{G}}{\partial \gamma_{gh}}\boldsymbol{G}^{-1}\tilde{\boldsymbol{u}}^{(w)}\right. \\
&\quad \left. - tr\left(\boldsymbol{G}^{-1}\frac{\partial \boldsymbol{G}}{\partial \gamma_{gh}}\boldsymbol{G}^{-1}\boldsymbol{C}^{ZZ^{(w)}}\right)\right].
\end{aligned}$$

An estimate for $\boldsymbol{\beta}$ can be obtained at the end of the REML EM algorithm by using the generalised least squares estimate of $\boldsymbol{\beta}$. Although we have presented the random effects approach to implementing a REML EM algorithm it is worth noting that Cullis, Smith & Thompson (2004) and Knight (2008) derived exactly the same parameter updates by considering $\boldsymbol{\beta}$ as fixed. The difference between their fixed effects approach and the random effects approach was that in the E-step the conditional expectation is with respect to $\boldsymbol{y}_2$ rather than $\boldsymbol{y}$.

3.2. A new incomplete data specification for a REML EM algorithm for the linear mixed model

We consider a linear mixed model specified in the same way as in (1). We define a new complete data specification $\boldsymbol{y}_c^{(U)} = (\boldsymbol{y}_2^\top, \boldsymbol{u}^\top)^\top$. This differs from the previously published specification $\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}^\top, \boldsymbol{\beta}^\top, \boldsymbol{u}^\top)^\top$ by considering the incomplete data to be the transformed observed data vector associated with the marginal log density function in (6) which is used for REML estimation of the variance parameters. To form the complete data log density function requires the joint distribution of $\boldsymbol{y}_2$ and $\boldsymbol{u}$, i.e.,

$$\begin{pmatrix} \boldsymbol{y}_2 \\ \boldsymbol{u} \end{pmatrix} \sim \mathrm{N}\left\{\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{L}_2^\top\boldsymbol{H}\boldsymbol{L}_2 & \boldsymbol{L}_2^\top\boldsymbol{Z}\boldsymbol{G} \\ \boldsymbol{G}\boldsymbol{Z}^\top\boldsymbol{L}_2 & \boldsymbol{G} \end{pmatrix}\right\}.$$

The complete data log density function using $\boldsymbol{y}_c^{(U)}$ can be written (excluding constants) as

$$\begin{aligned}
\log\{f(\boldsymbol{y}_c^{(U)}; \boldsymbol{\kappa})\} = -\frac{1}{2}\Big[&\log\{\det(\boldsymbol{L}_2^\top\boldsymbol{R}\boldsymbol{L}_2)\} + \log\{\det(\boldsymbol{G})\} + \boldsymbol{u}^\top\boldsymbol{G}^{-1}\boldsymbol{u} \\
&+ (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{u})^\top\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{u})\Big].
\end{aligned} \tag{9}$$

The relationship between the complete data log density functions based on $\boldsymbol{y}_c^{(U)}$ and $\boldsymbol{y}_c^{(P)}$ can be shown to be

$$\log\{f(\boldsymbol{y}_c^{(P)}; \boldsymbol{\theta})\} = \log\{f(\boldsymbol{y}_c^{(U)}; \boldsymbol{\kappa})\} + \log\{f(\boldsymbol{y}_1 | \boldsymbol{y}_2; \boldsymbol{\theta})\},$$

Table 1 Summary of REML EM algorithm complete data specifications for the linear mixed model presented in (1).

Complete data	$\boldsymbol{\beta}$ fixed or random	References
$\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}^\top, \boldsymbol{\beta}^\top, \boldsymbol{u}^\top)^\top$	random	Dempster, Laird & Rubin (1977); Laird & Ware (1982); Foulley, Jaffrezic & Robert-Cranie (2000)
$\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}^\top, \boldsymbol{u}^\top)^\top$	fixed	Knight (2008); Cullis, Smith & Thompson (2004)
$\boldsymbol{y}_c^{(U)} = (\boldsymbol{y}_2^\top, \boldsymbol{u}^\top)^\top$	fixed	Not published previously as far as we are aware

which is similar to the relation presented in (4) in the Verbyla (1990) conditional derivation of REML. This suggests that using the complete data specification $\boldsymbol{y}_c^{(P)}$ should be similar to using $\ell(\boldsymbol{\theta}; \boldsymbol{L}^\top \boldsymbol{y})$ for the REML estimation of variance parameters.

For the complete data specification $\boldsymbol{y}_c^{(U)}$ the E-step involves taking the conditional expectation of the complete data log density function in (9) with respect to $\boldsymbol{y}_2$ and at the current iterate of $\boldsymbol{\kappa}$, denoted $\boldsymbol{\kappa}^{(w)}$, i.e.,

$$\begin{aligned} Q^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w)}) &= E\left[\log\{f(\boldsymbol{y}_c^{(U)}; \boldsymbol{\kappa})\} | \boldsymbol{y}_2; \boldsymbol{\kappa}^{(w)} \right] \\ &= E\left[\log\{f(\boldsymbol{L}_2^\top \boldsymbol{e}; \sigma_e^2, \boldsymbol{\phi})\} | \boldsymbol{y}_2; \boldsymbol{\kappa}^{(w)} \right] + E\left[\log\{f(\boldsymbol{u}; \boldsymbol{\gamma})\} | \boldsymbol{y}_2; \boldsymbol{\kappa}^{(w)} \right] \\ &= Q^{(U)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\kappa}^{(w)}) + Q^{(U)}(\boldsymbol{\gamma}; \boldsymbol{\kappa}^{(w)}). \end{aligned}$$

Computation of $Q^{(U)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\kappa}^{(w)})$ and $Q^{(U)}(\boldsymbol{\gamma}; \boldsymbol{\kappa}^{(w)})$ only requires the conditional distribution $\boldsymbol{u}|\boldsymbol{y}_2$ which is equivalent to $\boldsymbol{u}|\boldsymbol{y}$ when applying the random effects approach to $\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}^\top, \boldsymbol{\beta}^\top, \boldsymbol{u}^\top)^\top$. We can write

$$\begin{aligned} Q^{(U)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\kappa}^{(w)}) &= -\frac{1}{2}\Big[\log\{\det(\boldsymbol{L}_2^\top \boldsymbol{R} \boldsymbol{L}_2)\} + (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)})^\top \boldsymbol{S} (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}) \\ &\qquad + tr\left(\boldsymbol{Z}^\top \boldsymbol{S} \boldsymbol{Z} \boldsymbol{C}^{\boldsymbol{Z}\boldsymbol{Z}^{(w)}} \right) \Big], \\ Q^{(U)}(\boldsymbol{\gamma}; \boldsymbol{\kappa}^{(w)}) &= Q^{(P)}(\boldsymbol{\gamma}; \boldsymbol{\theta}^{(w)}). \end{aligned}$$

The M-step of a REML EM algorithm for linear mixed models using $\boldsymbol{y}_c^{(U)}$ involves maximising $Q^{(U)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\kappa}^{(w)})$ and $Q^{(U)}(\boldsymbol{\gamma}; \boldsymbol{\kappa}^{(w)})$ with respect to $\sigma_e^2, \boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ respectively.

The updating equation for σ_e^2 is

$$\sigma_e^{2(w+1)} = \frac{1}{n-t}\left[(\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)})^\top \boldsymbol{U}^{(w)} (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}) + tr\left(\boldsymbol{Z}^\top \boldsymbol{U}^{(w)} \boldsymbol{Z} \boldsymbol{C}^{\boldsymbol{Z}\boldsymbol{Z}^{(w)}} \right) \right].$$

where $\boldsymbol{U}^{(w)} = \boldsymbol{\Sigma}^{-1^{(w)}} - \boldsymbol{\Sigma}^{-1^{(w)}} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1^{(w)}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1^{(w)}}$. The updating equation for ϕ_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n$, $i \geqslant j$ can be found by equating the following partial derivative to zero,

$$\begin{aligned} \frac{\partial Q^{(U)}(\sigma_e^2, \boldsymbol{\phi}; \boldsymbol{\kappa}^{(w)})}{\partial \phi_{ij}} &= -\frac{1}{2}\Big[tr\left(\boldsymbol{U} \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_{ij}} \right) - \frac{1}{\sigma_e^2} (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)})^\top \boldsymbol{U} \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_{ij}} \boldsymbol{U} (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}) \\ &\qquad - \frac{1}{\sigma_e^2} tr\left(\boldsymbol{Z}^\top \boldsymbol{U} \frac{\partial \boldsymbol{\Sigma}}{\partial \phi_{ij}} \boldsymbol{U} \boldsymbol{Z} \boldsymbol{C}^{\boldsymbol{Z}\boldsymbol{Z}^{(w)}} \right) \Big]. \end{aligned}$$

A summary of complete data specifications for a REML EM algorithm is provided in Table 1.

It is worth noting that firstly, although $\boldsymbol{y}_2$ constitutes the incomplete data of $\boldsymbol{y}_c^{(U)}$ it does not need to be explicitly computed. Secondly, the observed information matrix can be obtained using the method of Louis (1982) or Oakes (1999).

3.3. Outline of a proof that using $\boldsymbol{y}_c^{(U)} = (\boldsymbol{y}_2^\top; \boldsymbol{u}^\top)^\top$ will converge to a (local) maximum of the residual log-likelihood function

The proof can be obtained by following the steps in the example of Section 3 in Dempster, Laird & Rubin (1977). We begin by defining $g(\boldsymbol{y}_2; \boldsymbol{\kappa})$ as the marginal density function of $\boldsymbol{y}_2$ and $k(\boldsymbol{u}|\boldsymbol{y}_2; \boldsymbol{\kappa})$ as the conditional density function of $\boldsymbol{u}$ given $\boldsymbol{y}_2$ so that we can write

$$\log\{f(\boldsymbol{y}_c^{(U)}; \boldsymbol{\kappa})\} = \log\{g(\boldsymbol{y}_2; \boldsymbol{\kappa})\} + \log\{k(\boldsymbol{u}|\boldsymbol{y}_2; \boldsymbol{\kappa})\}. \tag{10}$$

Rearranging and taking the conditional expectation of both sides of (10) with respect to $\boldsymbol{y}_2$ and at the current iterate of $\boldsymbol{\kappa}$, denoted $\boldsymbol{\kappa}^{(w)}$, i.e., the E-step of the EM algorithm, we have

$$\ell(\boldsymbol{\kappa}; \boldsymbol{y}_2) = Q^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w)}) - H^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w)}),$$

where $Q^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w)}) = E[\log\{f(\boldsymbol{y}_c^{(U)}; \boldsymbol{\kappa})\}|\boldsymbol{y}_2; \boldsymbol{\kappa}^{(w)}]$ and $H^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w)}) = E[\log\{k(\boldsymbol{u}|\boldsymbol{y}_2; \boldsymbol{\kappa})\}|\boldsymbol{y}_2; \boldsymbol{\kappa}^{(w)}]$. Choosing an update of $\boldsymbol{\kappa}^{(w)}$, denoted $\boldsymbol{\kappa}^{(w+1)}$, such that $Q^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w+1)}) \geqslant Q^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w)})$, i.e., the M-step of the EM algorithm, will result in convergence to a (local) maximum of the residual log-likelihood function since the difference $H^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w+1)}) - H^{(U)}(\boldsymbol{\kappa}; \boldsymbol{\kappa}^{(w)})$ can be bounded by Jensen's inequality.

4. A new REML PX-EM algorithm for linear mixed models

Practical implementations imply that a REML PX-EM algorithm is to be preferred to a REML EM algorithm. The REML PX-EM algorithm, like the REML EM algorithm, has monotonic convergence and parameter updates remain in the parameter space. Furthermore, the E-step of a REML PX-EM algorithm is exactly the same as the E-step of a REML EM algorithm. The gains in using the PX-EM algorithm compared to the EM algorithm are in the rates of convergence. Liu, Rubin & Wu (1998) show that the PX-EM algorithm has a rate of convergence that is at least as fast as the EM algorithm, and often much faster.

Foulley & Van Dyk (2000) considered a REML PX-EM algorithm for linear mixed models using the random effects approach, i.e. the approach where $\boldsymbol{y}$ comprises the incomplete data and where the vector of fixed effects $\boldsymbol{\beta}$ is considered a random effect with variance tending to infinity. However a REML PX-EM algorithm where $\boldsymbol{y}_2$ constitutes the incomplete data has not been considered previously.

4.1. Linear mixed model formulation when using a REML PX-EM algorithm

For the REML PX-EM algorithm the linear mixed model is reformulated by introducing an auxiliary parameter $\boldsymbol{\lambda}$ (say) so that the linear mixed model in (1) is expanded to

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{f} + \boldsymbol{e}, \tag{11}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n \times t$ and $n \times b$ design matrices respectively, and $\Lambda = \Lambda(\boldsymbol{\lambda})$ is a $b \times b$ real invertible matrix which is a function of the $v \times 1$ auxiliary parameter vector $\boldsymbol{\lambda}$. It is assumed that the joint distribution of $\boldsymbol{f}$ and $\boldsymbol{\epsilon}$ is

$$\begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{e} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_* \end{pmatrix} \right\},$$

where $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{d})$ and $\boldsymbol{R}_* = \sigma^2_{e*}\boldsymbol{\Sigma}_*(\boldsymbol{\phi}_*)$ are symmetric positive definite matrices. The marginal distribution of $\boldsymbol{y}$ under the parameter expanded model is

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{H}_*),$$

where $\boldsymbol{H}_* = \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{D}\Lambda^\top\boldsymbol{Z}^\top + \boldsymbol{R}_*$. The expanded parameter set for the model in (11) is $\mathcal{K} = (\boldsymbol{\kappa}_*^\top, \boldsymbol{\lambda}^\top)^\top$ where $\boldsymbol{\kappa}_* = (\sigma^2_{e*}, \boldsymbol{d}^\top, \boldsymbol{\phi}_*^\top)^\top$. The subscript $*$ is used to distinguish variance parameters associated with the linear mixed model specified in (11) and the linear mixed model specified in (1). The expanded parameter vector $\mathcal{K}$ must satisfy two conditions:

- $\mathcal{K}$ can be reduced to the original parameter vector $\boldsymbol{\kappa} = (\sigma^2_e, \boldsymbol{\gamma}^\top, \boldsymbol{\phi}^\top)^\top$ by a many-to-one reduction function. For the model specified in (11) this reduction function, denoted by $\mathcal{R}$ is $\boldsymbol{\kappa} = (\sigma^2_e, \boldsymbol{\gamma}^\top, \boldsymbol{\phi}^\top)^\top = \mathcal{R}(\mathcal{K}) = (\sigma^2_{e*}, \mathrm{vech}(\boldsymbol{\Lambda}\boldsymbol{D}\Lambda^\top), \boldsymbol{\phi}_*^\top)^\top$.
- When the auxiliary parameter $\boldsymbol{\lambda}$ is set to its null value $\boldsymbol{\lambda}_0$ (say) the expanded complete data model is reduced to the original complete data model. If we define $\boldsymbol{u} = \boldsymbol{\Lambda}\boldsymbol{f}$ then $\boldsymbol{G} = \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}^\top$ and when $\boldsymbol{\Lambda}(\boldsymbol{\lambda}_0) = \boldsymbol{I}_b$ the expanded model in (11) is reduced to the original model in (1) with $\boldsymbol{u} = \boldsymbol{f}$ and $\boldsymbol{G} = \boldsymbol{D}$.

Liu, Rubin & Wu (1998) note that the idea of parameter expansion is to perform a covariance adjustment between the imputed missing data and the known incomplete data vector. In the context of a REML PX-EM algorithm for a linear mixed model the auxiliary parameter $\boldsymbol{\lambda}$ is used to "correct" the estimate of $\boldsymbol{G}$ to produce an adjusted estimate. At convergence there is no longer any need to adjust the estimate of $\boldsymbol{G}$ and $\boldsymbol{\Lambda}(\boldsymbol{\lambda}) = \boldsymbol{I}_b$. Another way to consider the auxiliary parameter in the REML PX-EM algorithm is that it performs a rotation on the vector of random effects in an attempt to orthogonalise the space in which variance parameters are estimated.

We have used a notation similar to that of Liu, Rubin & Wu (1998), i.e., the subscript $*$ identifies variance parameters in the parameter expanded model. However it is worth noting that $\boldsymbol{\phi}_* = \boldsymbol{\phi}$ and $\sigma^2_{e*} = \sigma^2_e$.

4.2. REML PX-EM algorithm using y_2 as the incomplete data

For a REML PX-EM algorithm for the linear mixed model we consider the complete data specification $\boldsymbol{y}_{c_\mathcal{X}}^{(U)} = (\boldsymbol{y}_2^\top, \boldsymbol{f}^\top)^\top$ where the subscript $\mathcal{X}$ is used to distinguish between a REML PX-EM algorithm and a REML EM algorithm. For the parameter expanded model based on $\boldsymbol{y}_{c_\mathcal{X}}^{(U)}$ the joint distribution of the incomplete and missing data is

$$\begin{pmatrix} \boldsymbol{y}_2 \\ \boldsymbol{f} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{L}_2^\top\boldsymbol{H}_*\boldsymbol{L}_2 & \boldsymbol{L}_2^\top\boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{\Lambda}^\top\boldsymbol{Z}^\top\boldsymbol{L}_2 & \boldsymbol{D} \end{pmatrix} \right\}, \tag{12}$$

where $\boldsymbol{H}_* = \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}^\top\boldsymbol{Z}^\top + \boldsymbol{R}_*$. The complete data log density function for $\boldsymbol{y}_c^{(U)}$ is

$$\begin{aligned} \log\{f_\mathcal{X}(\boldsymbol{y}_c^{(U)}; \mathcal{K})\} = -\frac{1}{2}\Big[&\log\{\det(\boldsymbol{L}_2^\top\boldsymbol{R}_*\boldsymbol{L}_2)\} + \log\{\det(\boldsymbol{D})\} + \boldsymbol{f}^\top\boldsymbol{D}^{-1}\boldsymbol{f} \\ &+ (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{f})^\top\boldsymbol{S}_*(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{f})\Big], \end{aligned} \tag{13}$$

where $\boldsymbol{S}_* = \boldsymbol{R}_*^{-1} - \boldsymbol{R}_*^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{R}_*^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{R}_*^{-1}$. The E-step of the REML PX-EM algorithm involves taking the conditional expectation of the complete data log density function in (13) with respect to $\boldsymbol{y}_2$ and at the current iterate of $\boldsymbol{\mathcal{K}}^{(w)} = \{\boldsymbol{\kappa}_*^{(w)^\top}, (\boldsymbol{\lambda} = \boldsymbol{\lambda}_0)^\top\}^\top$. Setting the auxiliary parameter equal to its null value at each iteration reduces the computations in the REML PX-EM algorithm E-step to those of the REML EM algorithm E-step. We can write

$$\begin{aligned} Q_{\mathcal{X}}^{(U)}(\boldsymbol{\mathcal{K}};\boldsymbol{\mathcal{K}}^{(w)}) &= E\left[\log\{f_{\mathcal{X}}(\boldsymbol{y}_c^{(U)};\boldsymbol{\mathcal{K}})\}|\boldsymbol{y}_2;\boldsymbol{\mathcal{K}}^{(w)}\right] \\ &= E\left[\log\{f_{\mathcal{X}}(\boldsymbol{L}_2^\top\boldsymbol{\epsilon};\boldsymbol{\mathcal{K}})\}|\boldsymbol{y}_2;\boldsymbol{\mathcal{K}}^{(w)}\right] + E\left[\log\{f_{\mathcal{X}}(\boldsymbol{f};\boldsymbol{d})\}|\boldsymbol{y}_2;\boldsymbol{\mathcal{K}}^{(w)}\right] \\ &= Q_{\mathcal{X}}^{(U)}(\sigma_{e*}^2,\boldsymbol{\phi}_*,\boldsymbol{\lambda};\boldsymbol{\mathcal{K}}^{(w)}) + Q_{\mathcal{X}}^{(U)}(\boldsymbol{d};\boldsymbol{\mathcal{K}}^{(w)}), \end{aligned}$$

where

$$\begin{aligned} Q_{\mathcal{X}}^{(U)}(\sigma_{e*}^2,\boldsymbol{\phi}_*,\boldsymbol{\lambda};\boldsymbol{\mathcal{K}}^{(w)}) &= -\frac{1}{2}\Big[\log\{\det(\boldsymbol{L}_2^\top\boldsymbol{R}_*\boldsymbol{L}_2)\} \\ &\qquad + E\Big\{(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{f})^\top\boldsymbol{S}_*(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\Lambda}f)|\boldsymbol{y}_2;\boldsymbol{\mathcal{K}}^{(w)}\Big\}\Big], \\ Q_{\mathcal{X}}^{(U)}(\boldsymbol{d};\boldsymbol{\mathcal{K}}^{(w)}) &= -\frac{1}{2}\Big[\log\{\det(\boldsymbol{D})\} + E\big(\boldsymbol{f}^\top\boldsymbol{D}^{-1}\boldsymbol{f}|\boldsymbol{y}_2;\boldsymbol{\mathcal{K}}^{(w)}\big)\Big]. \end{aligned}$$

The M-step of the REML PX-EM algorithm involves maximising $Q_{\mathcal{X}}^{(U)}(\boldsymbol{\mathcal{K}};\boldsymbol{\mathcal{K}}^{(w)})$ with respect to $\boldsymbol{\mathcal{K}} = (\sigma_{e*}^2, \boldsymbol{d}^\top, \boldsymbol{\phi}_*^\top, \boldsymbol{\lambda}^\top)^\top$ and then applying the reduction function $\mathcal{R}(\boldsymbol{\mathcal{K}}) = (\sigma_{e*}^2, \text{vech}(\boldsymbol{\Lambda}D\boldsymbol{\Lambda}^\top), \boldsymbol{\phi}_*^\top)^\top$, where vech refers to vector-half (see Section 12.9 in Searle 1982),to obtain estimates of $\boldsymbol{\kappa} = (\sigma_e^2, \boldsymbol{\gamma}^\top, \boldsymbol{\phi}^\top)^\top$. Regardless of the form of $\boldsymbol{R}_*$ and $\boldsymbol{D}$ an update for the auxiliary parameter $\boldsymbol{\lambda}$ can be written

$$\boldsymbol{\lambda}^{(w+1)} = \boldsymbol{A}^{-1^{(w)}}\boldsymbol{b}^{(w)},$$

where the elements of the $b^2 \times b^2$ matrix $\boldsymbol{A}^{(w)}$ are

$$a_{gh,qr}^{(w)} = \tilde{\boldsymbol{u}}^{(w)^\top}\boldsymbol{J}_{qr}^\top\boldsymbol{Z}^\top\boldsymbol{S}^{(w)}\boldsymbol{Z}\frac{\partial\boldsymbol{\Lambda}^{(w)}}{\partial\lambda_{gh}^{(w)}}\tilde{\boldsymbol{u}}^{(w)} + tr\left(\boldsymbol{J}_{qr}^\top\boldsymbol{Z}^\top\boldsymbol{S}^{(w)}\boldsymbol{Z}\frac{\partial\boldsymbol{\Lambda}^{(w)}}{\partial\lambda_{gh}^{(w)}}\boldsymbol{C}^{ZZ^{(w)}}\right),$$

where g, h, q, and r are $1,\ldots,b$ and $\boldsymbol{J}_{qr}$ is a $b \times b$ indicator matrix with a unit entry in row q, column r and zeros elsewhere. The elements of the $b^2 \times 1$ vector $\boldsymbol{b}^{(w)}$ are

$$b_{gh}^{(w)} = \boldsymbol{y}^\top\boldsymbol{S}^{(w)}\boldsymbol{Z}\frac{\partial\boldsymbol{\Lambda}^{(w)}}{\partial\lambda_{gh}^{(w)}}\tilde{\boldsymbol{u}}^{(w)}.$$

5. Results

We consider two examples: a sire model for lamb weight data and a balanced incomplete block soybean variety trial. We have chosen these two examples because the data are easily obtained and because it is relatively straight-forward to use a statistical computing environment such as R (R Core Team 2017) to implement the REML EM and REML PX-EM algorithms presented below.

For both examples we applied a REML EM algorithm using the complete data specifications $\boldsymbol{y}_c^{(P)}$ and $\boldsymbol{y}_c^{(U)}$, and a REML PX-EM algorithm using the complete data specifications $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$ and $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$. We do not present the parameter updating equations for the REML EM algorithms

Table 2 Number of iterations to achieve convergence for the REML EM algorithms based on the complete data specifications $\boldsymbol{y}_c^{(P)}$ and $\boldsymbol{y}_c^{(U)}$ and REML PX-EM algorithms based on $\boldsymbol{y}_{c_\mathcal{X}}^{(P)}$ and $\boldsymbol{y}_{c_\mathcal{X}}^{(U)}$ applied to the lamb weight and soybean variety trial yield data.

Algorithm	Spec.	Iter. (SV1)	Iter. (SV2)
Sire model for lamb weight data			
REML EM	$\boldsymbol{y}_c^{(P)}$	339	341
REML EM	$\boldsymbol{y}_c^{(U)}$	339	340
REML PX-EM	$\boldsymbol{y}_{c_\mathcal{X}}^{(P)}$	76	77
REML PX-EM	$\boldsymbol{y}_{c_\mathcal{X}}^{(U)}$	54	54
AI		11	fail
Soybean variety trial yield data			
REML EM	$\boldsymbol{y}_c^{(P)}$	18	18
REML EM	$\boldsymbol{y}_c^{(U)}$	14	16
REML PX-EM	$\boldsymbol{y}_{c_\mathcal{X}}^{(P)}$	17	18
REML PX-EM	$\boldsymbol{y}_{c_\mathcal{X}}^{(U)}$	12	13
AI		8	fail

Note: The first set of starting values (SV1) for the lamb weight data were $\sigma_e^{2(0)} = 2$ and $\sigma_u^{2(0)} = 2$ and for the soybean variety trial yield data were $\sigma_e^{2(0)} = 1$ and $\sigma_u^{2(0)} = 1$. The second set of starting values (SV2) for the lamb weight data were $\sigma_e^{2(0)} = 2$ and $\sigma_u^{2(0)} = 3$ and for the soybean variety trial yield data were $\sigma_e^{2(0)} = 8$ and $\sigma_u^{2(0)} = 4$. The abbreviations "Spec." and "Iter." mean specification and iterations respectively.

using the complete data specifications $\boldsymbol{y}_c^{(P)}$ and $\boldsymbol{y}_c^{(U)}$ as they can be obtained by making the substitution $\boldsymbol{\Lambda} = \boldsymbol{I}_b$ in the updating equations for the REML PX-EM algorithms based on $\boldsymbol{y}_{c_\mathcal{X}}^{(P)}$ and $\boldsymbol{y}_{c_\mathcal{X}}^{(U)}$ respectively. We also considered the AI algorithm described by Gilmour, Thompson & Cullis (1995) which is a popular Newton-Raphson type algorithm for fitting these types of models. We have used the `ASReml-R` (Version 3) software package (Butler 2009) within the `R` computing environment to implement the AI algorithm. The standard implementation of the AI algorithm within the `ASReml-R` software package involves a rescaling of the AI matrix when computing parameter updates. To implement the algorithm as described by Gilmour, Thompson & Cullis (1995), i.e., without a rescaling of the AI matrix, the `ASReml-R` package was used to obtain the AI matrix and the score for a given set of parameter estimates and parameter updates were computed manually within the `R` computing environment. The `R` code and data files for reproducing the convergence information in Table 2 are available at www.github.com/SimonDiffey/ANZJS_REMLPXEM.

In all cases we used the convergence criterion proposed by Foulley, Jaffrezic & Robert-Cranie (2000), i.e., the criterion that algorithm is stopped when

$$\sqrt{(\boldsymbol{\kappa}^{(w+1)} - \boldsymbol{\kappa}^{(w)})^\top (\boldsymbol{\kappa}^{(w+1)} - \boldsymbol{\kappa}^{(w)}) / \boldsymbol{\kappa}^{(w)\top} \boldsymbol{\kappa}^{(w)}} < 10^{-8},$$

where $\boldsymbol{\kappa}$ is the vector of variance parameters.

5.1. Example: lamb weight data

We analysed the lamb weight data presented in Callanan & Harville (1991) which consists of birth weights (in pounds) of single birth male lambs that are the progeny of 62

ewes. The ages of the ewes are categorised into three groups. The lambs were sired by one of 23 rams that belonged to five different population lines. The number of observations per sire is highly unbalanced. At one extreme there are seven sires that have a single observation each and at the other extreme there is a single sire that has nine observations.

A possible way of modelling these data would be to fit sire as a random effect and the overall mean, ewe age and line as fixed effects. We define $\boldsymbol{y}$ as a 62×1 vector of observed birth weights and $\boldsymbol{X}=(\boldsymbol{1}, \boldsymbol{X}_a, \boldsymbol{X}_\ell)$ as a 62×7 design matrix of full column rank parameterised with the corner point constraint. The term $\boldsymbol{1}$ is a 62×1 vector of 1s. The design matrices $\boldsymbol{X}_a$ and $\boldsymbol{X}_l$ are indicator matrices for the second and third age groups and the second to the fifth population line respectively. The vector of fixed effects is similarly partitioned as $\boldsymbol{\beta}=(\mu, \boldsymbol{\beta}_a^\top, \boldsymbol{\beta}_l^\top)^\top$. We assumed that $\boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_{23})$, the vector of rescaled random effects is $\boldsymbol{f} \sim \mathrm{N}(\boldsymbol{0}, d\boldsymbol{I}_{23})$, and the 62×23 matrix $\boldsymbol{Z}$ is their associated design matrix. It is assumed that the vector of residuals is distributed $\boldsymbol{\epsilon} \sim \mathrm{N}(\boldsymbol{0}, \sigma_{e*}^2 \boldsymbol{I}_{62})$. The invertible matrix associated with the working parameter is $\boldsymbol{\Lambda}=\lambda \boldsymbol{I}_{23}$ where λ is scalar. We define the variance parameter vector of interest to be $\boldsymbol{\kappa}=(\sigma_e^2, \sigma_u^2)^\top$ and the expanded parameter vector to be $\boldsymbol{\mathcal{K}}=(\sigma_{e*}^2, d, \lambda)^\top$.

The updating equations for a REML PX-EM algorithm based on the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}=(\boldsymbol{y}^\top, \boldsymbol{\beta}^\top, \boldsymbol{f}^\top)^\top$ are

$$
\begin{aligned}
d^{(w+1)} &= \frac{1}{b}\left\{ \tilde{\boldsymbol{u}}^{(w)^\top} \tilde{\boldsymbol{u}}^{(w)} + tr\left(\boldsymbol{C}^{\mathcal{Z}\mathcal{Z}^{(w)}}\right) \right\}, \\
\sigma_{e*}^{2^{(w+1)}} &= \frac{1}{n}\left\{ \tilde{\boldsymbol{e}}^{(w)^\top} \tilde{\boldsymbol{e}}^{(w)} + tr\left(\boldsymbol{W}\boldsymbol{C}^{-1^{(w)}} \boldsymbol{W}^\top\right) \right\}, \\
\lambda_P^{(w+1)} &= \frac{\tilde{\boldsymbol{u}}^{(w)^\top} \boldsymbol{Z}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(w)}) - tr\left(\boldsymbol{Z}^\top \boldsymbol{X} \boldsymbol{C}^{\mathcal{X}\mathcal{Z}^{(w)}}\right)}{\tilde{\boldsymbol{u}}^{(w)^\top} \boldsymbol{Z}^\top \boldsymbol{Z} \tilde{\boldsymbol{u}}^{(w)} + tr\left(\boldsymbol{Z}^\top \boldsymbol{Z} \boldsymbol{C}^{\mathcal{Z}\mathcal{Z}^{(w)}}\right)},
\end{aligned}
$$

where we use the subscript P to distinguish between the update for λ based on the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$ and $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$. Note that $b=23$, $n=62$, and $\boldsymbol{\beta}^{(w)}$ is the generalised least squares estimate of $\boldsymbol{\beta}$ at the w-th iterate. The updates for σ_{e*}^2 can be re-expressed as

$$
\sigma_{e*}^{2^{(w+1)}} = \frac{1}{n}\left\{ (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)})^\top \boldsymbol{K} (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}) + tr\left(\boldsymbol{Z}^\top \boldsymbol{K} \boldsymbol{Z} \boldsymbol{C}^{\mathcal{Z}\mathcal{Z}^{(w)}}\right) + t\sigma_{e*}^{2^{(w)}} \right\} \tag{14}
$$

where $\boldsymbol{K}=\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$ and $t=7$. When using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}=(\boldsymbol{y}_2^\top, \boldsymbol{f}^\top)^\top$ the updating equation for d is the same as that employed when using $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$. The updating equations for the other parameters are

$$
\begin{aligned}
\sigma_{e*}^{2^{(w+1)}} &= \frac{1}{n-t}\left\{ (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)})^\top \boldsymbol{K} (\boldsymbol{y} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}) + tr\left(\boldsymbol{Z}^\top \boldsymbol{K} \boldsymbol{Z} \boldsymbol{C}^{\mathcal{Z}\mathcal{Z}^{(w)}}\right) \right\}, \qquad (15) \\
\lambda_U^{(w+1)} &= \frac{\boldsymbol{y}^\top \boldsymbol{K} \boldsymbol{Z} \tilde{\boldsymbol{u}}^{(w)}}{\tilde{\boldsymbol{u}}^{(w)^\top} \boldsymbol{Z}^\top \boldsymbol{K} \boldsymbol{Z} \tilde{\boldsymbol{u}}^{(w)} + tr\left(\boldsymbol{Z}^\top \boldsymbol{K} \boldsymbol{Z} \boldsymbol{C}^{\mathcal{Z}\mathcal{Z}^{(w)}}\right)}.
\end{aligned}
$$

The difference between (14) and (15) is subtle but illuminating. In the latter the degrees of freedom associated with the estimation of the fixed effects are incorporated directly into the divisor, whereas in the former they are incorporated in the term $t\sigma_{e*}^{2^{(w)}}$. In the early iterations of a REML (PX)EM algorithm (14) and (15) will result in different updates although the two will be equivalent at convergence.

In regards to the relationship between λ_U and λ_P it can be shown, using identities arising from the mixed model equations and the fact that $tr(\boldsymbol{C}\boldsymbol{C}^{-1}) = \boldsymbol{I}_{t+b}$, that

$$\lambda_U^{(w+1)} - 1 = \frac{a}{b}(\lambda_P^{(w+1)} - 1), \tag{16}$$

i.e., the left hand side of (16) is a rescaling of the right hand side where $a = \tilde{\boldsymbol{u}}^{(w)^\top}\boldsymbol{Z}^\top\boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)} + tr\left(\boldsymbol{Z}^\top\boldsymbol{Z}\boldsymbol{C}^{ZZ^{(w)}}\right)$ and $b = \tilde{\boldsymbol{u}}^{(w)^\top}\boldsymbol{Z}^\top\boldsymbol{K}\boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)} + tr\left(\boldsymbol{Z}^\top\boldsymbol{K}\boldsymbol{Z}\boldsymbol{C}^{ZZ^{(w)}}\right)$.

We used two sets of starting values for σ_e^2 and σ_u^2. The first consisted of $\sigma_e^{2^{(0)}} = 2$ and $\sigma_u^{2^{(0)}} = 2$, and the second of $\sigma_e^{2^{(0)}} = 2$ and $\sigma_u^{2^{(0)}} = 3$. Using the updating equations above, applying the reduction function $\boldsymbol{\kappa} = \mathcal{R}(\boldsymbol{\mathcal{K}}) = (\sigma_{e*}^2, d\lambda^2)^\top$, and applying the convergence criteria to the parameter vector $\boldsymbol{\kappa} = (\sigma_e^2, \sigma_u^2)^\top$, we obtained REML estimates of the variance parameters σ_e^2 and σ_u^2 given by $\hat{\sigma}_e^2 = 2.961$ and $\hat{\sigma}_u^2 = 0.517$. The number of iterations to achieve convergence for all algorithms applied to the lamb weight data are presented in Table 2.

5.2. Example: a balanced incomplete block soybean variety trial

The soybean variety trial data originate from table 1 of a technical report by Weiss & Cox (1939). The experimental design is a balanced incomplete block design consisting of 6 replicate plots of 31 soybean varieties and 31 blocks. The plots were 6 feet by 16 feet and each contained two rows of soybeans.

We fitted a model where the overall mean and the effect of soybean variety were taken to be fixed effects. Block was fitted as a random effect. We define $\boldsymbol{y}$ to be a 186×1 vector of observed yields (bushels per acre) and $\boldsymbol{X} = (\boldsymbol{1}, \boldsymbol{X}_v)$ as a 186×31 matrix of full column rank parameterised with the corner point constraint. The design matrix $\boldsymbol{X}_v$ is an indicator matrix for variety number two through to variety number thirty one. The vector of fixed effects is $\boldsymbol{\beta} = (\mu, \boldsymbol{\beta}_v^\top)^\top$ and we assume that $\boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \sigma_u^2\boldsymbol{I}_{31})$ where σ_u^2 is a variance component associated with between block variation. We also assume that $\boldsymbol{e} \sim \mathrm{N}(\boldsymbol{0}, \sigma_e^2\boldsymbol{I}_{186})$ and that $\boldsymbol{u}$ and $\boldsymbol{e}$ are independent of each other.

For the REML PX-EM algorithms we set the vector of rescaled random effects associated with blocks equal to $\boldsymbol{f} \sim \mathrm{N}(\boldsymbol{0}, d\boldsymbol{I}_{31})$ and we assume that the vector of residuals is $\boldsymbol{\epsilon} \sim \mathrm{N}(\boldsymbol{0}, \sigma_{e*}^2\boldsymbol{I}_{186})$. The invertible matrix associated with the working parameter is $\boldsymbol{\Lambda} = \lambda\boldsymbol{I}_{31}$. We define the variance parameter vector of interest to be $\boldsymbol{\kappa} = (\sigma_e^2, \sigma_u^2)^\top$ and the expanded parameter vector to be $\boldsymbol{\mathcal{K}} = (\sigma_{e*}^2, d, \lambda)^\top$.

The updating equations for the REML PX-EM algorithm based on the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)} = (\boldsymbol{y}^\top, \boldsymbol{\beta}^\top, \boldsymbol{f}_a^\top)^\top$ and $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)} = (\boldsymbol{y}_2^\top, \boldsymbol{\beta}^\top, \boldsymbol{f}_a^\top)^\top$ have the same form as those presented for the lamb weight data and we note that $n = 186$, $t = 31$ and $b = 31$.

We used two sets of starting values for σ_e^2 and σ_u^2. The first consisted of $\sigma_e^{2^{(0)}} = 1$ and $\sigma_u^{2^{(0)}} = 1$. The second consisted of $\sigma_e^{2^{(0)}} = 8$ and $\sigma_u^{2^{(0)}} = 4$. Using the updating equations, applying the reduction function $\boldsymbol{\kappa} = \mathcal{R}(\boldsymbol{\mathcal{K}}) = (\sigma_{e*}^2, d\lambda^2)^\top$, and applying the convergence criteria to the parameter vector $\boldsymbol{\kappa} = (\sigma_e^2, \sigma_u^2)^\top$, we found the REML estimates of the variance parameters σ_e^2 and σ_u^2 to be $\hat{\sigma}_e^2 = 3.585$ and $\hat{\sigma}_u^2 = 5.267$. The number of iterations to achieve convergence for all algorithms applied to the soybean variety trial data are presented in Table 2.

6. Discussion

The AI algorithm is a Newton-Raphson type algorithm and in cases where the algorithm converges it will outperform REML EM algorithms and their variants. However, a problem

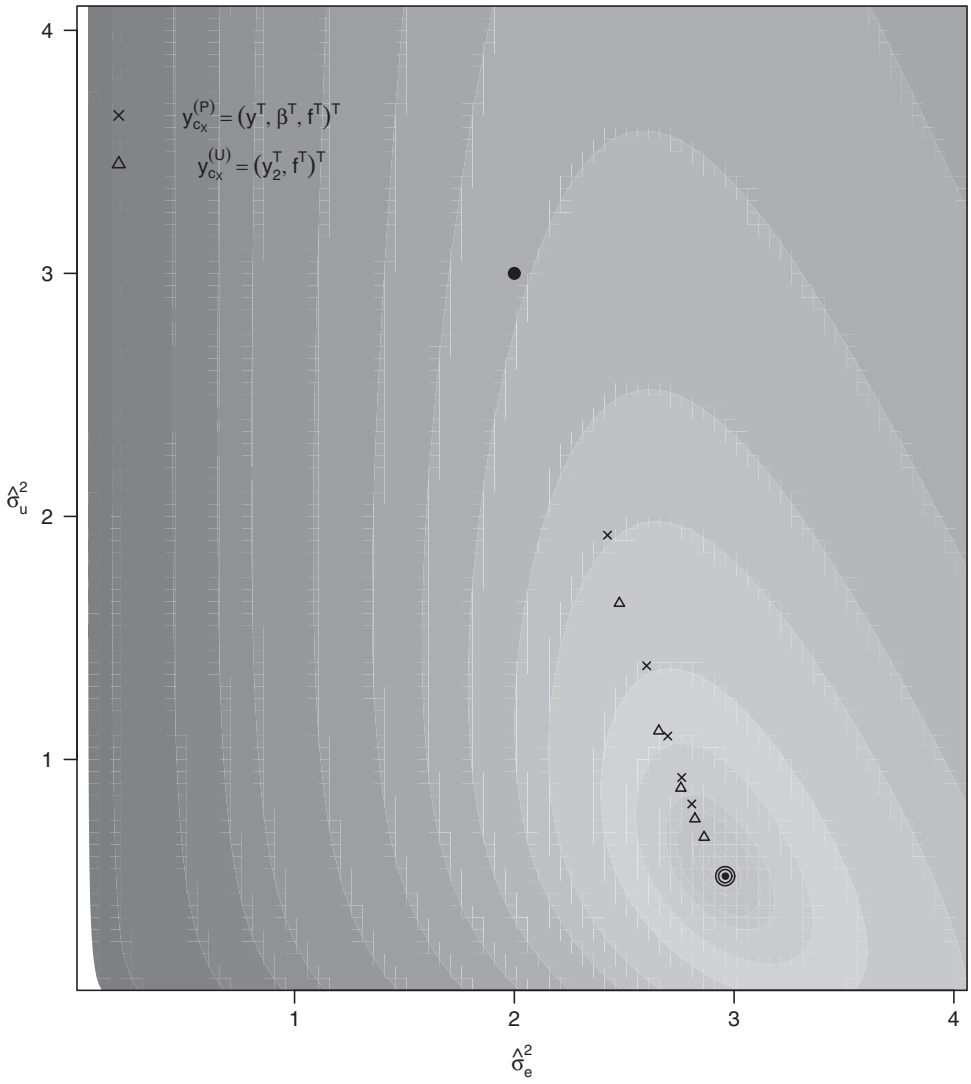

Figure 1. Contour plot of the residual log-likelihood surface for the lamb weight data and the first 5 iterations of the REML PX-EM algorithms using the complete data specification $\mathbf{y}_{c\mathcal{X}}^{(U)}$ and $\mathbf{y}_{c\mathcal{X}}^{(P)}$. The starting values of $\sigma_e^{2(0)} = 2$ and $\sigma_u^{2(0)} = 3$ are represented by the black circle. The REML estimates of $\hat{\sigma}_e^2 = 2.962$ and $\hat{\sigma}_u^2 = 0.517$ are represented by the "target".

with Newton-Raphson type algorithms is that the variance parameter updates are not guaranteed to remain in the parameter space and the algorithm can fail to converge. For the sire model for the lamb weight data and the incomplete block model for the soybean variety yield data, we considered a set of starting values where this is the case (Table 2). REML EM algorithms and their variants have the desirable properties of monotonic convergence and parameter updates remaining in the parameter space. On the other hand they can be slow to converge. In Table 2 we considered the performance of four variants of the REML EM algorithm.

We introduce a new incomplete data specification, referred to as $\boldsymbol{y}_2$ which is based on the transformed observed data vector associated with the marginal distribution in the conditional derivation of REML (Verbyla 1990). This is a useful alternative to the current practice of considering the incomplete data to be $\boldsymbol{y}$. For both the lamb weight and the soybean variety trial data the number of iterations to convergence was less when using the incomplete data specification $\boldsymbol{y}_2$ compared to $\boldsymbol{y}$. There was only one instance where the number of iterations to convergence was the same for the two alternative REML EM algorithms (Table 2). In the soybean variety trial example a REML EM algorithm using the new incomplete data specification converged in fewer iterations than a REML PX-EM algorithm using the current practice of considering the incomplete data to be $\boldsymbol{y}$. A REML PX-EM algorithm where the incomplete data are defined as $\boldsymbol{y}_2$ resulted in fewer iterations to convergence for both the lamb weight and soybean variety trial data. Depending on starting values, convergence was achieved in 28% to 30% fewer iterations in both examples. We have conducted simulation studies based on a one-way random effects model and in more than 99% of cases a REML PX-EM algorithm based on specifying the incomplete data as $\boldsymbol{y}_2$ converged in fewer iterations than a REML PX-EM algorithm using $\boldsymbol{y}$ as the incomplete data.

A contour plot of the REML log-likelihood surface for the lamb weight data and the first five iterations of the REML PX-EM algorithms using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$ and $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$ with the starting values $\sigma_e^{2(0)} = 2$ and $\sigma_u^{2(0)} = 3$ is provided in Figure 1. With this set of starting values the AI algorithm fails. There are two outstanding features of Figure 1; firstly, the REML PX-EM algorithm based on $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$ comes closer to the solution after every iterate than does a REML PX-EM algorithm based on $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$ and secondly, a REML PX-EM algorithm based on $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$ is close to the solution after a small number of iterations. This second feature suggests that a hybrid AI and REML PX-EM algorithm might be of some benefit in certain cases. Where the AI algorithm fails an alternative might be to consider a hybrid algorithm that starts with a small number of REML PX-EM algorithm iterations before switching to the AI algorithm. This is an approach considered by Meyer (2006), Knight (2008) and Zhou & Stephens (2014).

References

BUTLER, D. (2009). asreml: asreml() fits the linear mixed model. Available from URL: www.vsni.co.uk. R package version 3.0.

CALLANAN, T.P. & HARVILLE, D.A. (1991). Some new algorithms for computing restricted maximum likelihood estimates of variance components. *Journal of Statistical Computation and Simulation* **38**, 239–259.

CULLIS, B., SMITH, A. & THOMPSON, R. (2004). Perspectives of ANOVA, REML, and a general linear mixed model. In *Methods and Models in Statistics*, eds. N. Adams, M. Crowder, D.J. Hand & D. Stephens, pp. 53–94. London: Imperial College Press. In honour of Professor John Nelder, FRS.

DEMPSTER, A., LAIRD, N. & RUBIN, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.

FOULLEY, J.L., JAFFREZIC, F. & ROBERT-CRANIE, C. (2000). EM-REML estimation of covariance parameters in Gaussian mixed models for longitudinal data analysis. *Genetics Selection Evolution* **32**, 129–141.

FOULLEY, J.L. & VAN DYK, D. (2000). The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genetics Selection Evolution* **32**, 143–163.

GILMOUR, A., THOMPSON, R. & CULLIS, B. (1995). Average Information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440–1450.

HENDERSON, C., KEMPTHORNE, O., SEARLE, S. & VON KROSIGK, C. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218.

KNIGHT, E. (2008). Improved iterative schemes for REML estimation of variance parameters in linear mixed models. Ph.D. thesis, The University of Adelaide, School of Agriculture, Food and Wine.

LAIRD, N. & WARE, J. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

LIU, C., RUBIN, D. & WU, Y. (1998). Parameter expansion to accelerate the EM algorithm: the PX-EM algorithm. *Biometrika* **85**, 755–770.

LOUIS, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society*, *B* **44**, 226–233.

MEYER, K. (2006). PX × AI: Algorithmics for better convergence in restricted maximum likelihood estimation. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brasil. Available from URL: http://didgeridoo.une.edu.au/km/StuffToDownload/8WCGALP/pxai.pdf. Last downloaded 28/08/2017.

OAKES, D. (1999). *Direct calculation of the information matrix via the EM algorithm. Journal of the Royal Statistical Society*, B, **61**, 479–482.

PATTERSON, H. & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from URL: https://www.R-project.org/.

SEARLE, S.R. (1982). *Matrix Algebra Useful for Statistics*. Wiley Series in Probability and Mathematical Statistics: Applied probability and Statistics. New York, NY: Wiley.

SEARLE, S., CASELLA, G. & MCCULLOCH, C. (2006). *Variance Components*. Hoboken, New Jersey: John Wiley & Sons.

SMYTH, G. (1998). Optimization and Nonlinear Equations. In *Encyclopedia of Biostatistics*, eds. P. Armitage and T. Colton, pp. 3174–3180. Chichester: John Wiley & Sons.

VERBYLA, A. (1990). A conditional derivation of residual maximum likelihood. *Australian Journal of Statistics* **32**, 227–230.

WEISS, M.G. & COX, G.M. (1939). Balanced incomplete block and lattice square designs for testing yield differences among large numbers of soybean varieties. *Research Bulletin (Iowa Agriculture and Home Economics Experiment Station)* **23**, 1.

ZHOU, X. & STEPHENS, M. (2014). Efficient algorithms for multivariate linear mixed models in genomewide association studies. *Nature Methods* **11**, 407.