NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

2-13

A faster and computationally more efficient REML (PX)EM algorithm for linear mixed models

Simon Diffey, Alan Wesh, Alison Smith and Brian Cullis

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

A faster and computationally more efficient REML (PX)EM algorithm for linear mixed models

Simon Diffey[1,*], Alan Welsh[2], Alison Smith[1], and Brian Cullis[1]

[1]National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, NSW, Australia
[2]Centre for Mathematics and its Applications, Australian National University, Canberra, ACT, Australia
*email: sdiffey@uow.edu.au

SUMMARY: Residual maximum likelihood is the preferred method for estimating variance parameters associated with a linear mixed model. Typically an iterative algorithm is required for the estimation of these parameters. Two algorithms which can be used for this purpose are the EM algorithm and the PX-EM algorithm. Both require specification of the complete data which comprises the incomplete and missing data. We consider a new incomplete data specification which is computationally more efficient than alternative specifications. In the example considered the new incomplete data specification results in the algorithm converging in 30% fewer iterations than the alternative specification. We describe the conditions necessary for this faster rate of convergence to apply in other cases.

KEY WORDS: EM algorithm; PX-EM algorithm; Linear mixed model; Variance components; Residual maximum likelihood.

1. Introduction

Linear mixed models are routinely applied to biological data arising from designed experiments. The preferred method for estimating the parameters associated with these models is residual maximum likelihood (REML) (Patterson and 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 average information algorithm (Gilmour, Thompson, and 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 may result in the algorithm failing to converge to a solution.

The EM algorithm (Dempster, Laird, and Rubin, 1977) and the parameter expanded EM (PX-EM) algorithm (Liu, Rubin, and Wu, 1998) have the desirable properties of monotonic convergence and parameter updates remaining in their parameter space. Dempster et al. (1977) and Liu et al. (1998) only consider maximum likelihood estimation although REML estimation is briefly mentioned by Dempster et al. (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 and Ware (1982) and Foulley, Jafrezic, and Robert-Cranie (2000). A REML PX-EM algorithm for linear mixed models was considered by Foulley and 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 et al. (1977).

Before the REML EM or REML PX–EM algorithm can be considered practical alternatives to Newton-Raphson type algorithms two issues need to be addressed. Firstly, they can be notoriously slow to converge, particularly the REML EM algorithm. Secondly, compared to the average information algorithm, current implementations of these two algorithms are computationally more expensive at each iterate. This increased computational expense relates to calculating the trace of a matrix of the same order as the length of the observed data vector.

The purpose of this paper is to present a new incomplete data specification of a REML EM and REML PX–EM algorithm for linear mixed models. This new incomplete data specification is computationally more efficient and we describe the conditions under which this new specification will have a faster rate of convergence than current specifications.

This paper proceeds in four parts. First, we provide a statistical background which consists of a linear mixed model formulation, 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. This consists of the complete data specification used in current implementations of a REML EM algorithm for linear mixed models and a new implementation based on a new incomplete data specification. We compare the computational requirements of the new implementation to current implementations. We sketch a proof that this new implementation will converge to a (local) maximum of the residual log–likelihood function and present the conditions necessary for the new implementation to have a faster rate of convergence. Third, we consider a REML PX–EM algorithm for linear mixed models using the new implementation. Fourth, we illustrate the use of the current and new implementation of both a REML EM and REML

PX–EM algorithm for linear mixed models by considering an example.

2. Statistical background

2.1 *Linear mixed model formulation*

The linear mixed model can be written as

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

where $\boldsymbol{y}_o$ is a $n \times 1$ vector of observed data, $\boldsymbol{\beta}$ is a $t \times 1$ vector of fixed effects and $\boldsymbol{X}$ is its associated design matrix and 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 assumed positive definite and depends on the parameter vector $\boldsymbol{\gamma}$. It is assumed that $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{R})$ where $\boldsymbol{R} = \sigma^2\boldsymbol{\Sigma}(\boldsymbol{\phi})$ and the matrix $\boldsymbol{\Sigma}(\boldsymbol{\phi})$ is assumed positive definite and depends on the parameter vector $\boldsymbol{\phi}$. The joint distribution of $\boldsymbol{y}_o$, $\boldsymbol{u}$, and $\boldsymbol{e}$ is

$$\begin{pmatrix} \boldsymbol{y}_o \\ \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{X}\boldsymbol{\beta} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{H} & \boldsymbol{Z}\boldsymbol{G} & \boldsymbol{R} \\ \boldsymbol{G}\boldsymbol{Z}^T & \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{R} & \boldsymbol{0} & \boldsymbol{R} \end{pmatrix} \right\} \tag{2}$$

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

2.2 *Henderson's mixed model equations*

Dempster et al. (1977) in Section 4.4 of their paper provide an example of an EM algorithm for a linear mixed model. They note that

> in the context of linear models...the relevance of incomplete-data concepts is at first sight remote.

We believe Henderson's mixed model equations is a natural place to start when considering an EM algorithm for the linear mixed model. A maximum likelihood (ML) EM algorithm for a linear mixed model and Henderson's mixed model equations are both derived by considering the same log-density function. Furthermore, we use the coefficient matrix associated with Henderson's mixed model equations and its inverse to compare the computational requirements of alternative REML EM and REML PX–EM algorithms.

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

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

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

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

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

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

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

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

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

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

and the inverse of the coefficient matrix as

$$\boldsymbol{C}^{-1} = \begin{pmatrix} \boldsymbol{C}^{XX} & \boldsymbol{C}^{XZ} \\ \boldsymbol{C}^{ZX} & \boldsymbol{C}^{ZZ} \end{pmatrix}$$

where $\boldsymbol{C}^{XX} = (\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}$, $\boldsymbol{C}^{XZ} = \boldsymbol{C}^{ZX^T} = -(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{Z}\boldsymbol{G}$, $\boldsymbol{C}^{ZZ} = (\boldsymbol{Z}^T\boldsymbol{S}\boldsymbol{Z} + \boldsymbol{G}^{-1})^{-1}$, and $\boldsymbol{S} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{R}^{-1}$. 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's mixed model equations.

2.3 *A conditional derivation of REML*

The preferred method for estimating the variance parameters associated with a linear mixed model is residual or restricted maximum likelihood (REML). The original reference for REML is the paper by Patterson and 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 in Verbyla (1990). The conditional derivation of REML begins by considering the transformation

$$\boldsymbol{L}^T\boldsymbol{y}_o = \begin{pmatrix} \boldsymbol{L}_1^T\boldsymbol{y}_o \\ \boldsymbol{L}_2^T\boldsymbol{y}_o \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix}$$

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 $\boldsymbol{L}_1^T\boldsymbol{X} = \boldsymbol{I}_t$ and $\boldsymbol{L}_2^T\boldsymbol{X} = \boldsymbol{0}$. The distribution of the transformed data is

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

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 $\boldsymbol{L}^T\boldsymbol{y}_o$ can be expressed as

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

Since the vector $\boldsymbol{y}_1$ is of length t 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. Therefore, 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

$$\begin{aligned}\ell(\boldsymbol{\kappa};\boldsymbol{y}_2) = -\tfrac{1}{2}\big[&(n-t)\log(2\pi) + \log\{\det(\boldsymbol{L}^T\boldsymbol{L})\} \\ &+ \log\{\det(\boldsymbol{H})\} + \log\{\det(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})\} + \boldsymbol{y}_o^T\boldsymbol{P}\boldsymbol{y}_o\big] \quad (5)\end{aligned}$$

where $\boldsymbol{P} = \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}$. Ignoring the constant $\log\{\det(\boldsymbol{L}^T\boldsymbol{L})\}$ in (5) results in the form of the residual log-likelihood function most commonly presented.

3. Two alternative complete data specifications

3.1 *Current implementations of a REML EM algorithm for the linear mixed model*

Current implementations of REML EM and REML PX–EM algorithms for linear mixed models consider 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 consider $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \boldsymbol{B})$ and the complete data to be defined as $\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}_o^T, \boldsymbol{\beta}^T, \boldsymbol{u}^T)^T$ where the superscript P on $\boldsymbol{y}_c$ will be used to distinguish this complete data specification from the new complete data specification presented shortly. The random effects approach is the approach briefly described by Dempster et al. (1977) and followed by others including Laird and Ware (1982), Foulley et al. (2000), and Foulley and Van Dyk (2000). We define the joint distribution of $\boldsymbol{y}_o$, $\boldsymbol{\beta}$, $\boldsymbol{u}$, and $\boldsymbol{e}$ as

$$\begin{pmatrix} \boldsymbol{y}_o \\ \boldsymbol{\beta} \\ \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{F} & \boldsymbol{XB} & \boldsymbol{ZG} & \boldsymbol{R} \\ \boldsymbol{BX}^T & \boldsymbol{B} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{GZ}^T & \boldsymbol{0} & \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{R} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R} \end{pmatrix} \right\}, \quad (6)$$

where

$$\begin{aligned}\boldsymbol{F} &= \boldsymbol{XBX}^T + \boldsymbol{ZGZ}^T + \boldsymbol{R} \\ &= \boldsymbol{XBX}^T + \boldsymbol{H}\end{aligned}$$

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

$$\begin{aligned}\boldsymbol{F}^{-1} &\to \boldsymbol{H}^{-1} - \boldsymbol{H}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{H}^{-1}, \\ \boldsymbol{F}^{-1} &\to \boldsymbol{P}.\end{aligned}$$

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

$$\begin{aligned}\log\{f(\boldsymbol{y}_c;\boldsymbol{\theta})\} &= -\frac{1}{2}\Big[\log\{\det(\boldsymbol{R})\} + \log\{\det(\boldsymbol{G})\} \\ &\quad + \boldsymbol{u}^T\boldsymbol{G}^{-1}\boldsymbol{u} + \boldsymbol{e}^T\boldsymbol{R}^{-1}\boldsymbol{e}\Big], \quad (7)\end{aligned}$$

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

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 (7) with respect to $\boldsymbol{y}_o$ and at the current iterate of $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}^{(w)}$, i.e.,

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

Computation of $Q^{(P)}(\sigma^2,\boldsymbol{\phi};\boldsymbol{\theta}^{(w)})$ and $Q^{(P)}(\boldsymbol{\gamma};\boldsymbol{\theta}^{(w)})$ requires the conditional distributions $\boldsymbol{e}|\boldsymbol{y}_o$ and $\boldsymbol{u}|\boldsymbol{y}_o$ respectively. Using (6) it can be shown that

$$\begin{aligned}\boldsymbol{e}|\boldsymbol{y}_o &\sim N\left(\boldsymbol{RP}\boldsymbol{y}_o, \boldsymbol{WC}^{-1}\boldsymbol{W}^T\right) \\ \boldsymbol{u}|\boldsymbol{y}_o &\sim N\left(\boldsymbol{GZ}^T\boldsymbol{P}\boldsymbol{y}_o, \boldsymbol{C}^{ZZ}\right)\end{aligned}$$

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

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

where $\tilde{\boldsymbol{e}}^{(w)} = \boldsymbol{R}^{(w)}\boldsymbol{P}^{(w)}\boldsymbol{y}_o$ and $\tilde{\boldsymbol{u}}^{(w)} = \boldsymbol{G}^{(w)}\boldsymbol{Z}^T\boldsymbol{P}^{(w)}\boldsymbol{y}_o$. A distinguishing feature of the E–step associated with using the complete data specification $\boldsymbol{y}_c^{(P)}$ is that $Q^{(P)}(\sigma^2,\boldsymbol{\phi};\boldsymbol{\theta}^{(w)})$ requires calculating the trace of $n \times n$ matrix which involves the inverse of the full coefficient matrix associated with Henderson's mixed model equations.

The M-step of a REML EM algorithm for linear mixed models using $\boldsymbol{y}_c^{(P)}$ involves maximising $Q^{(P)}(\sigma^2,\boldsymbol{\phi};\boldsymbol{\theta}^{(w)})$ and $Q^{(P)}(\boldsymbol{\gamma};\boldsymbol{\theta}^{(w)})$ with respect to $\sigma^2, \boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ respectively. 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 et al. (2004) and Knight, E. (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 $\boldsymbol{y}_o$.

3.2 *A new implementation of a REML EM algorithm for the linear mixed model*

We consider a linear mixed model specified in the same way as in Section 2.1. We define a new complete data specification $\boldsymbol{y}_c^{(U)} = (\boldsymbol{y}_2^T, \boldsymbol{u}^T)^T$. This differs from the previously published specification $\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}_o^T, \boldsymbol{\beta}^T, \boldsymbol{u}^T)^T$ by considering the incomplete data to be the transformed observed data vector

associated with the marginal log–density function in (5) which is used for the REML estimation of variance parameters. To form the complete data log density function requires the joint distribution of $\boldsymbol{y}_2$ and $\boldsymbol{u}$, i.e.,

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

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

$$\begin{aligned} \log\{f(\boldsymbol{y}_c^{(U)};\boldsymbol{\kappa})\} &= -\frac{1}{2}\Big[\log\{\det(\boldsymbol{L}_2^T\boldsymbol{R}\boldsymbol{L}_2)\} + \log\{\det(\boldsymbol{G})\} \\ &\quad +(\boldsymbol{y}_o - \boldsymbol{Z}\boldsymbol{u})^T\boldsymbol{S}(\boldsymbol{y}_o - \boldsymbol{Z}\boldsymbol{u}) + \boldsymbol{u}^T\boldsymbol{G}^{-1}\boldsymbol{u}\Big]. \end{aligned} \tag{8}$$

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

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

which is similar to the relation presented in (4) based on the Verbyla (1990) conditional derivation of REML. This suggests that using the complete data specification $\boldsymbol{y}_c^{(P)}$ would be similar to using $\ell(\boldsymbol{\theta};\boldsymbol{L}^T\boldsymbol{y}_o)$ for the REML estimation of variance parameters. While it is possible to use both complete data specifications for the REML estimation of variance parameters the task is simplified when considering the complete data specification $\boldsymbol{y}_c^{(U)}$. The advantages of using $\boldsymbol{y}_c^{(U)}$ compared to $\boldsymbol{y}_c^{(P)}$ are more apparent when considering the E–step of a REML EM algorithm based on the complete data specification $\boldsymbol{y}_c^{(U)}$. The E–step involves taking the conditional expectation of the complete data log density function in (8) with respect to $\boldsymbol{y}_2$ and at the current iterate of $\boldsymbol{\kappa}$, denoted $\boldsymbol{\kappa}^{(w)}$, i.e.,

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

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

$$\begin{aligned} Q^{(U)}(\sigma^2,\boldsymbol{\phi};\boldsymbol{\kappa}^{(w)}) &= -\frac{1}{2}\Big[\log\{\det(\boldsymbol{L}_2^T\boldsymbol{R}\boldsymbol{L}_2)\} \\ &\quad +(\boldsymbol{y}_o - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)})^T\boldsymbol{S}(\boldsymbol{y}_o - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}) \\ &\quad +\mathrm{tr}\big(\boldsymbol{Z}^T\boldsymbol{S}\boldsymbol{Z}\boldsymbol{C}^{ZZ^{(w)}}\big)\Big], \\ Q^{(U)}(\boldsymbol{\gamma};\boldsymbol{\kappa}^{(w)}) &= Q^{(P)}(\boldsymbol{\gamma};\boldsymbol{\theta}^{(w)}). \end{aligned}$$

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

It is worth noting that although $\boldsymbol{y}_2$ comprises the incomplete data of $\boldsymbol{y}_c^{(U)}$ it does not need to be explicitly formed. In regard to the E–step there are two advantages to using $\boldsymbol{y}_c^{(U)}$ compared to $\boldsymbol{y}_c^{(P)}$. Firstly, only one conditional distribution is required, namely $\boldsymbol{u}|\boldsymbol{y}_2$, when using $\boldsymbol{y}_c^{(U)}$. Secondly, $\boldsymbol{y}_c^{(U)}$ only requires calculating traces of matrices of order b. These involve only a subset of the inverse of the coefficient matrix associated with Henderson's mixed model equations, i.e., $\boldsymbol{C}^{ZZ}$. The complete data specification $\boldsymbol{y}_c^{(P)}$ requires calculating the trace of a matrix of order n which involves the full inverse of the coefficient matrix associated with Henderson's mixed model equations. The E–step of a REML EM algorithm for linear mixed models is the same as the E–step of a REML PX–EM algorithm. Therefore, either algorithm where the incomplete data is considered to be $\boldsymbol{y}_2$ will be computationally more efficient. This increase in computational efficiency depends on the size of n and the relative sizes of n and b.

3.3 *Sketch of a proof that using $\boldsymbol{y}_c^{(U)} = (\boldsymbol{y}_2^T,\boldsymbol{u}^T)^T$ will converge to a (local) maximum of the residual log–likelihood function*

We only sketch a proof here since a thorough derivation can be obtained by following the example of Dempster et al. (1977) Section 3. 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

$$\log\{f(\boldsymbol{y}_c^{(U)};\boldsymbol{\kappa})\} = \log\{g(\boldsymbol{y}_2;\boldsymbol{\kappa})\} + \log\{k(\boldsymbol{u}|\boldsymbol{y}_2;\boldsymbol{\kappa})\} \tag{9}$$

Rearranging and taking the conditional expectation of both sides of (9) 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 $H^{(U)}(\boldsymbol{\kappa};\boldsymbol{\kappa}^{(w)}) = E[\log\{k(\boldsymbol{u}|\boldsymbol{y}_2;\boldsymbol{\kappa})\}|\boldsymbol{y}_2;\boldsymbol{\kappa}^{(w)}]$. By 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 through the use of Jensen's inequality.

3.4 *Conditions necessary for a REML EM algorithm based on $\boldsymbol{y}_c^{(U)} = (\boldsymbol{y}_2^T,\boldsymbol{u}^T)^T$ to have a faster rate of convergence than $\boldsymbol{y}_c^{(P)} = (\boldsymbol{y}_o^T,\boldsymbol{\beta}^T,\boldsymbol{u}^T)^T$*

For ease of notation we define the complete data information matrices associated with the variance parameters for $\boldsymbol{y}_c^{(U)}$ and $\boldsymbol{y}_c^{(P)}$, and the observed information matrix as

$$\begin{aligned} \boldsymbol{\mathcal{I}}_c^{(U)} &= \boldsymbol{\mathcal{I}}_c^{(U)}(\boldsymbol{\kappa}^\star;\boldsymbol{y}_c^{(U)}) = -\left\{\frac{\partial^2 Q^{(U)}(\boldsymbol{\kappa};\boldsymbol{\kappa}^\star)}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^T}\right\}_{\boldsymbol{\kappa}=\boldsymbol{\kappa}^\star}, \\ \boldsymbol{\mathcal{I}}_c^{(P)} &= \boldsymbol{\mathcal{I}}_c^{(P)}(\boldsymbol{\kappa}^\star;\boldsymbol{y}_c^{(P)}) = -\left\{\frac{\partial^2 Q^{(P)}(\boldsymbol{\kappa};\boldsymbol{\kappa}^\star)}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^T}\right\}_{\boldsymbol{\kappa}=\boldsymbol{\kappa}^\star}, \\ \mathrm{I}_o &= \mathrm{I}_o(\boldsymbol{\kappa}^\star;\boldsymbol{y}_o), \end{aligned}$$

where $\boldsymbol{\kappa}^{(w)}$ has converged to $\boldsymbol{\kappa}^\star$. The observed information matrix can be calculated using the method of Louis (1982) or Oakes (1999). We define the rate matrices

$$\begin{aligned} \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star) &= \boldsymbol{I}_d - \boldsymbol{\mathcal{I}}_c^{(U)^{-1}}\mathrm{I}_o, \\ \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star) &= \boldsymbol{I}_d - \boldsymbol{\mathcal{I}}_c^{(P)^{-1}}\mathrm{I}_o, \end{aligned}$$

which can be written as

$$\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star) = {\boldsymbol{\mathcal{I}}_c^{(U)}}^{-1}\mathrm{I}_o,$$
$$\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star) = {\boldsymbol{\mathcal{I}}_c^{(P)}}^{-1}\mathrm{I}_o.$$

The rate of convergence associated with the complete data specification $\boldsymbol{y}_c^{(U)}$ and $\boldsymbol{y}_c^{(P)}$ is $\lambda_{\min}\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star)\}$ and $\lambda_{\min}\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star)\}$ respectively (McLachlan and Krishnan, 1997, Section 3.9.1). The term $\lambda_{\min}(.)$ refers to the minimum eigenvalue. We aim to show

$$\lambda_{\min}\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star)\} \geqslant \lambda_{\min}\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star)\}, \qquad (10)$$

i.e., the rate of convergence of a REML EM algorithm using the complete data specification $\boldsymbol{y}_c^{(U)}$ is faster than the rate of convergence of a REML EM algorithm using the complete data specification $\boldsymbol{y}_c^{(P)}$. We have not been able to prove that this inequality always holds. However, in our experience the rate of convergence when using $\boldsymbol{y}_c^{(U)}$ has always been faster than that when using $\boldsymbol{y}_c^{(P)}$. By considering a Loewner partial ordering (Horn and Johnson, 1990, Chapter 7) of the complete information matrices associated with $\boldsymbol{y}_c^{(U)}$ and $\boldsymbol{y}_c^{(P)}$ we are able to present the conditions under which the inequality in (10) will hold.

If we consider the difference between $\boldsymbol{\mathcal{I}}_c^{(P)}$ and $\boldsymbol{\mathcal{I}}_c^{(U)}$ to be positive definite then we can define the following Loewner partial ordering

$$\boldsymbol{\mathcal{I}}_c^{(U)} \preceq \boldsymbol{\mathcal{I}}_c^{(P)},$$

and as a consequence we can write

$${\boldsymbol{\mathcal{I}}_c^{(U)}}^{-1} \succeq {\boldsymbol{\mathcal{I}}_c^{(P)}}^{-1},$$

and

$$\det({\boldsymbol{\mathcal{I}}_c^{(U)}}^{-1}) \geqslant \det({\boldsymbol{\mathcal{I}}_c^{(P)}}^{-1}), \qquad (11)$$
$$\det(\boldsymbol{\mathcal{I}}_c^{(U)}) \leqslant \det(\boldsymbol{\mathcal{I}}_c^{(P)}), \qquad (12)$$

where going from (11) to (12) uses a result pertaining to the determinant of a matrix inverse. Multiplying both sides of (12) by $\det(\mathrm{I}_o)$ we have

$$\begin{aligned}
\det({\boldsymbol{\mathcal{I}}_c^{(U)}}^{-1})\det(\mathrm{I}_o) &\geqslant \det({\boldsymbol{\mathcal{I}}_c^{(P)}}^{-1})\det(\mathrm{I}_o),\\
\det({\boldsymbol{\mathcal{I}}_c^{(U)}}^{-1}\mathrm{I}_o) &\geqslant \det({\boldsymbol{\mathcal{I}}_c^{(P)}}^{-1}\mathrm{I}_o),\\
\det\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star)\} &\geqslant \det\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star)\}.
\end{aligned}$$

It follows that

$$\boldsymbol{\lambda}_k\big\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star)\big\} \geqslant \boldsymbol{\lambda}_k\big\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star)\big\},$$

where $\boldsymbol{\lambda}_k(.)$ is a vector of k eigenvalues. If the k eigenvalues of $\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star)$ and $\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star)$ are arranged in descending order then

$$\lambda_{\min}\big\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(U)}(\boldsymbol{\kappa}^\star)\big\} \geqslant \lambda_{\min}\big\{\boldsymbol{I}_d - \boldsymbol{\mathcal{J}}^{(P)}(\boldsymbol{\kappa}^\star)\big\},$$

which concludes the proof. The key part of the proof is the inequality in (12). Elements of the complete data information matrices for both $\boldsymbol{y}_c^{(U)}$ and $\boldsymbol{y}_c^{(P)}$ can be used to check if the difference between $\boldsymbol{\mathcal{I}}_c^{(P)}$ and $\boldsymbol{\mathcal{I}}_c^{(U)}$ is positive definite for particular cases.

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

From a practical implementation point of view a REML PX–EM algorithm is preferred to a REML EM algorithm as there is nothing to lose and often a lot to gain. Nothing to lose in the sense that the REML PX–EM algorithm, like the REML EM algorithm, has monotonic convergence and parameter updates remain in their 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 rates of convergence. Liu et al. (1998) show that the PX–EM algorithm has a rate of convergence that is no slower than the EM algorithm but is often much faster.

Foulley and Van Dyk (2000) considered a REML PX–EM algorithm for linear mixed models using the random effects approach, i.e., where $\boldsymbol{y}_o$ comprised 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 $\boldsymbol{y}_2$ comprises 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 formulated by introducing the auxiliary parameter $\boldsymbol{\lambda}$ (say) so that the linear mixed model in (1) is expanded to

$$\boldsymbol{y}_o = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{f} + \boldsymbol{\epsilon}, \qquad (13)$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $n\times t$ and $n\times b$ design matrices respectively, and $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\lambda})$ 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

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

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

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

where $\boldsymbol{H}_* = \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}^T\boldsymbol{Z}^T + \boldsymbol{R}_*$. The expanded parameter set for the model in (13) is $\boldsymbol{\mathcal{K}} = (\boldsymbol{\kappa}_*, \boldsymbol{\lambda}^T)^T$ where $\boldsymbol{\kappa}_* = (\sigma_*^2, \boldsymbol{d}^T, \boldsymbol{\phi}_*^T)^T$. The subscript $*$ is used to distinguish variance parameters associated with the linear mixed model specified in (13) and the linear mixed model specified in (1). The expanded parameter vector $\boldsymbol{\mathcal{K}}$ must satisfy two conditions:

(1) $\boldsymbol{\mathcal{K}}$ can be reduced to the original parameter vector $\boldsymbol{\kappa} = (\sigma^2, \boldsymbol{\gamma}^T, \boldsymbol{\phi}^T)^T$ by a many-to-one reduction function. For the model specified in (13) this reduction function, labelled $\mathcal{R}$ is $\boldsymbol{\kappa} = (\sigma^2, \boldsymbol{\gamma}^T, \boldsymbol{\phi}^T)^T = \mathcal{R}(\boldsymbol{\mathcal{K}}) = (\sigma_*^2, \mathrm{vech}(\boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}^T), \boldsymbol{\phi}_*^T)^T$.

(2) 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 $\boldsymbol{G} = \boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}^T$ and when $\boldsymbol{\Lambda}(\boldsymbol{\lambda}_0) = \boldsymbol{I}_b$ the expanded model in (13) is reduced to the original model in (1) with $\boldsymbol{u} = \boldsymbol{f}$ and $\boldsymbol{G} = \boldsymbol{D}$.

Liu et al. (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 $\boldsymbol{\Lambda}(\boldsymbol{\lambda}) = \boldsymbol{I}_b$.

4.2 *REML PX–EM algorithm using $\boldsymbol{y}_2$ as the incomplete data*

For a REML PX–EM algorithm for the linear mixed model we consider the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)} = (\boldsymbol{y}_2^T, \boldsymbol{f}^T)^T$ 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

$$\begin{pmatrix} \boldsymbol{y}_2 \\ \boldsymbol{f} \end{pmatrix} \sim N\left\{ \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{L}_2^T \boldsymbol{H}_* \boldsymbol{L}_2 & \boldsymbol{L}_2^T \boldsymbol{Z} \boldsymbol{\Lambda} \boldsymbol{D} \\ \boldsymbol{D} \boldsymbol{\Lambda}^T \boldsymbol{Z}^T \boldsymbol{L}_2 & \boldsymbol{D} \end{pmatrix} \right\} \quad (14)$$

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

$$\begin{aligned} \log\{f_{\mathcal{X}}(\boldsymbol{y}_c^{(U)}; \boldsymbol{\mathcal{K}})\} &= -\frac{1}{2}\Big[\log\{\det(\boldsymbol{L}_2^T \boldsymbol{R}_* \boldsymbol{L}_2)\} \\ &\quad + \log\{\det(\boldsymbol{D})\} + \boldsymbol{f}^T \boldsymbol{D}^{-1} \boldsymbol{f} \\ &\quad + (\boldsymbol{y}_o - \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{f})^T \boldsymbol{S}_* (\boldsymbol{y}_o - \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{f})\Big], \end{aligned} \quad (15)$$

where $\boldsymbol{S}_* = \boldsymbol{R}_*^{-1} - \boldsymbol{R}_*^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{R}_*^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{R}_*^{-1}$. The E–step of the REML PX–EM algorithm involves taking the conditional expectation of the complete data log density function in (15) with respect to $\boldsymbol{y}_2$ and at the current iterate of $\boldsymbol{\mathcal{K}}^{(w)} = \{\boldsymbol{\kappa}_*^{(w)^T}, (\boldsymbol{\lambda} = \boldsymbol{\lambda}_0)^T\}^T$. Setting the auxiliary parameter equal to its null value at each iterate reduces the computations in the REML PX–EM algorithm E–step to those of the REML EM algorithm E–step. We can write

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

where

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

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 $\boldsymbol{\mathcal{K}} = (\sigma_*^2, \boldsymbol{d}^T, \boldsymbol{\phi}_*^T, \boldsymbol{\lambda}^T)^T$ and then applying the reduction function $\mathcal{R}(\boldsymbol{\mathcal{K}}) = (\sigma_*^2, \text{vech}(\boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\Lambda}^T), \boldsymbol{\phi}_*^T)^T$ to obtain estimates of $\boldsymbol{\kappa} = (\sigma^2, \boldsymbol{\gamma}^T, \boldsymbol{\phi}^T)^T$. If we assume for simplicity that $\boldsymbol{R}_* = \sigma_*^2 \boldsymbol{I}_n$ then it can be shown that the parameter updates for $\boldsymbol{\mathcal{K}}$ are

$$\begin{aligned} \sigma_*^{2^{(w+1)}} &= \frac{1}{n-t}\Big\{(\boldsymbol{y}_o - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)})^T \boldsymbol{K} (\boldsymbol{y}_o - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}) \\ &\quad + \text{tr}\big(\boldsymbol{Z}^T \boldsymbol{K} \boldsymbol{Z} \boldsymbol{C}^{ZZ^{(w)}}\big)\Big\} \end{aligned}$$

where $\boldsymbol{K} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$.

$$\begin{aligned} d^{(w+1)} &= \frac{1}{b}\Big\{\tilde{\boldsymbol{u}}^{(w)^T} \tilde{\boldsymbol{u}}^{(w)} + \text{tr}\big(\boldsymbol{C}^{ZZ^{(w)}}\big)\Big\}, \\ \boldsymbol{\lambda}^{(w+1)} &= \boldsymbol{A}^{-1^{(w)}} \boldsymbol{b}^{(w)} \end{aligned}$$

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

$$\begin{aligned} a_{gh,qr}^{(w)} &= \tilde{\boldsymbol{u}}^{(w)^T} \boldsymbol{J}_{qr}^T \boldsymbol{Z}^T \boldsymbol{S}^{(w)} \boldsymbol{Z} \frac{\partial \boldsymbol{\Lambda}^{(w)}}{\partial \lambda_{gh}^{(w)}} \tilde{\boldsymbol{u}}^{(w)} \\ &\quad + \text{tr}\Big(\boldsymbol{J}_{qr}^T \boldsymbol{Z}^T \boldsymbol{S}^{(w)} \boldsymbol{Z} \frac{\partial \boldsymbol{\Lambda}^{(w)}}{\partial \lambda_{gh}^{(w)}} \boldsymbol{C}^{ZZ^{(w)}}\Big) \end{aligned}$$

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

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

5. Example: lamb weight data

We apply 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 only present the parameter updating equations associated with the REML PX–EM algorithms. The updating equations for the REML EM algorithms are a special case of these equations and can be obtained by making the substitution $\boldsymbol{\Lambda} = \boldsymbol{I}_b$. We use equivalent starting values for both the REML EM and REML PX–EM algorithms, however starting values are reported in terms relevant to the REML EM algorithm as we use an identity matrix of order b as the starting value for $\boldsymbol{\Lambda}$ in the REML PX–EM algorithms. We use the convergence criteria proposed by Foulley et al. (2000), i.e., the algorithm is stopped when

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

We consider the lamb weight data presented in Callanan and Harville (1991) which consists of birth weights (in pounds) of single birth male lambs that are the progeny of 62 ewes. The age of the ewes are categorised into 3 groups. The lambs were sired by one of 23 rams that belonged to 5 different population lines. A possible model for this data is to fit sire as a random effect and the overall mean, age and line as fixed effects. We define $\boldsymbol{y}_o$ as a 62×1 vector of observed birth weights, $\boldsymbol{X} = (\boldsymbol{1}\ \boldsymbol{X}_a\ \boldsymbol{X}_l)$ as a 62×7 design matrix of full column rank parameterized with the corner point constraint. The term $\boldsymbol{1}$ is a 62×1 vector of 1's. 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 $\boldsymbol{\beta} = (\mu, \boldsymbol{\beta}_a^T, \boldsymbol{\beta}_l^T)^T$. It is assumed that $\boldsymbol{u} \sim N(\mathbf{0}, \sigma_s^2 \boldsymbol{I}_{23})$, the vector of rescaled random effects is $\boldsymbol{f} \sim N(\mathbf{0}, d\boldsymbol{I}_{23})$, and the 62×23 matrix $\boldsymbol{Z}$ are their associated design matrix. It is assumed that the vector of residuals is distributed $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_*^2 \boldsymbol{I}_{62})$. 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 as $\boldsymbol{\kappa} = (\sigma^2, \sigma_s^2)^T$ and the expanded parameter vector as $\boldsymbol{\mathcal{K}} = (\sigma_*^2, d, \lambda)^T$. We consider two sets of starting values, $\boldsymbol{\kappa}^{(0)} = (1, 0.01)^T$ and $\boldsymbol{\kappa}^{(0)} = (1, 5)^T$. The updating equations for a REML PX–EM algorithm based on the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)} = (\boldsymbol{y}_o^T, \boldsymbol{\beta}^T, \boldsymbol{f}^T)^T$ are

$$
\begin{aligned}
d^{(w+1)} &= \frac{1}{23}\Big\{\tilde{\boldsymbol{u}}^{(w)^T}\tilde{\boldsymbol{u}}^{(w)} + \mathrm{tr}\big(\boldsymbol{C}^{ZZ^{(w)}}\big)\Big\} \\
\sigma_*^{2^{(w+1)}} &= \frac{1}{62}\Big\{\tilde{\boldsymbol{e}}^{(w)^T}\tilde{\boldsymbol{e}}^{(w)} + \mathrm{tr}\big(\boldsymbol{W}\boldsymbol{C}^{-1^{(w)}}\boldsymbol{W}^T\big)\Big\} \\
\lambda^{(w+1)} &= \frac{\tilde{\boldsymbol{u}}^{(w)^T}\boldsymbol{Z}^T(\boldsymbol{y}_o - \boldsymbol{X}\boldsymbol{\beta}^{(w)}) + \mathrm{tr}\big(\boldsymbol{Z}^T\boldsymbol{X}\boldsymbol{C}^{XZ^{(w)}}\big)}{\tilde{\boldsymbol{u}}^{(w)^T}\boldsymbol{Z}^T\boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)} + \mathrm{tr}\big(\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{C}^{ZZ^{(w)}}\big)}
\end{aligned}
$$

where $\boldsymbol{\beta}^{(w)}$ is the generalised least squares estimate of $\boldsymbol{\beta}$ at the w-th iterate. When using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)} = (\boldsymbol{y}_2^T, \boldsymbol{f}^T)^T$ the updating equation for d is the same as that when using $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$. The updating equations for the other parameters are

$$
\begin{aligned}
\sigma_*^{2^{(w+1)}} &= \frac{1}{55}\Big\{\big(\boldsymbol{y}_o - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}\big)^T\boldsymbol{K}\big(\boldsymbol{y}_o - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}\big) \\
&\quad + \mathrm{tr}\big(\boldsymbol{Z}^T\boldsymbol{K}\boldsymbol{Z}\boldsymbol{C}^{ZZ^{(w)}}\big)\Big\} \\
\lambda^{(w+1)} &= \frac{\boldsymbol{y}_o^T\boldsymbol{K}\boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)}}{\tilde{\boldsymbol{u}}^{(w)^T}\boldsymbol{Z}^T\boldsymbol{K}\boldsymbol{Z}\tilde{\boldsymbol{u}}^{(w)} + \mathrm{tr}\big(\boldsymbol{Z}^T\boldsymbol{K}\boldsymbol{Z}\boldsymbol{C}^{ZZ^{(w)}}\big)}
\end{aligned}
$$

Using these updating equations, the reduction function $\boldsymbol{\kappa} = \mathcal{R}(\boldsymbol{\mathcal{K}}) = (\sigma_*^2, d\lambda^2)^T$, and applying the convergence criteria to the parameter vector $\boldsymbol{\kappa} = (\sigma_s^2, \sigma^2)^T$, the REML estimates of the variance parameters σ^2 and σ_s^2 are $\hat{\sigma}^2 = 2.9616$ and $\hat{\sigma}_s^2 = 0.5171$.

Table 1

Iterations to convergence and rate of convergence (r) for two REML EM and REML PX–EM algorithms applied to the lamb weight data.

Algorithm	$\boldsymbol{y}_c$	Iter.	r
REML EM[a]	$\boldsymbol{y}_c^{(P)}$	1296	0.96307
REML EM[a]	$\boldsymbol{y}_c^{(U)}$	1296	0.96300
REML PX–EM[a]	$\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$	83	0.81679
REML PX–EM[a]	$\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$	57	0.74350
REML EM[b]	$\boldsymbol{y}_c^{(P)}$	342	0.96307
REML EM[b]	$\boldsymbol{y}_c^{(U)}$	341	0.96300
REML PX–EM[b]	$\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$	78	0.81679
REML PX–EM[b]	$\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$	55	0.74350

[a] $\boldsymbol{\kappa}^{(0)} = (1, 0.01)^T$
[b] $\boldsymbol{\kappa}^{(0)} = (1, 5)^T$

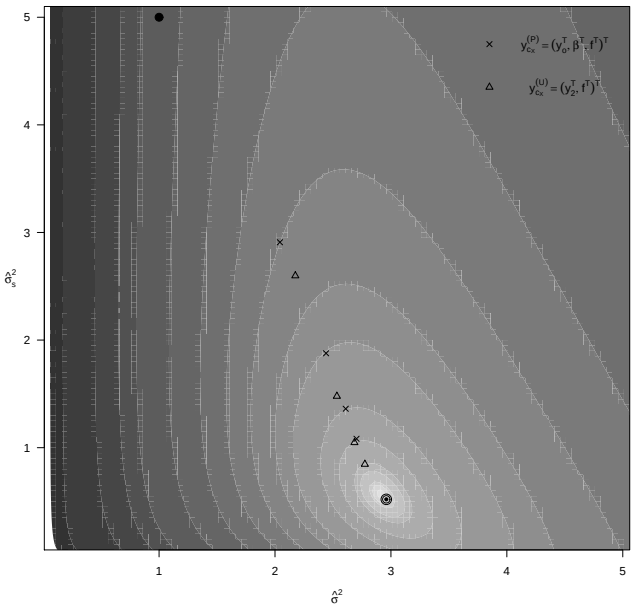

Figure 1. Contour plot of the residual log-likelihood surface for the lamb weight data and the first 4 iterations of the REML PX–EM algorithms using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$ and $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$. The starting values of $\sigma^{2^{(0)}} = 1$ and $\sigma_s^{2^{(0)}} = 5$ are represented by the black circle. The REML estimates of $\hat{\sigma}^2 = 2.962$ and $\hat{\sigma}_s^2 = 0.517$ are represented by the "target".

6. Discussion

The new incomplete data specification, referred to as $\boldsymbol{y}_2$ and based on the transformed observed data vector associated with the marginal distribution in the conditional derivation of REML (Verbyla, 1990) is a useful alternative to the incomplete data specification $\boldsymbol{y}_o$. Table 1 highlights the benefit of using a REML PX–EM algorithm compared to a REML EM algorithm and using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$ rather than $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$. In the lamb weight data case, using a REML PX–EM algorithm where the incomplete data is defined as $\boldsymbol{y}_2$ results in a faster rate of convergence and convergence is achieved in approximately 30% fewer iterations for both sets of starting values considered. A contour plot of the REML log–likelihood surface and the first four iterations of the REML PX–EM algorithms using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$ and $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$ is provided in Figure 1. This figure illustrates the faster rate of convergence associated with using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$. When using this complete data specification a larger initial step is taken towards the REML solution. By the third iterate a REML PX–EM using $\boldsymbol{y}_{c_{\mathcal{X}}}^{(U)}$ is approximately equivalent to the fourth iterate of the REML PX–EM algorithm using the complete data specification $\boldsymbol{y}_{c_{\mathcal{X}}}^{(P)}$. Figure 1 also highlights another feature of EM algorithms in general, which is that they often get close to a solution quite quickly but then take a long time to converge.

Our experience using a REML PX–EM based on specifying the incomplete data as $\boldsymbol{y}_2$ is that it will result in a faster rate

of convergence and converge in fewer iterations than a REML PX–EM algorithm where the incomplete data is specified as $\boldsymbol{y}_o$. We have used Loewner partial ordering to describe the conditions necessary for this to be the case. The other major benefit in using a REML EM or REML PX–EM algorithm where the incomplete data is specified as $\boldsymbol{y}_2$ is in regard to computational efficiency.

There are two related aspects to the improvement in computational efficiency when specifying the incomplete data as $\boldsymbol{y}_2$. Firstly, the E–step of a REML EM or REML PX–EM algorithm using the new incomplete data specification only requires knowledge of one conditional distribution, namely $\boldsymbol{u}|\boldsymbol{y}_2$. In contrast, the E–step of a REML EM and REML PX–EM algorithm using $\boldsymbol{y}_o$ as the incomplete data requires at least two conditional distributions, namely $\boldsymbol{u}|\boldsymbol{y}_2$ and typically $\boldsymbol{e}|\boldsymbol{y}_2$. Secondly, REML EM and REML PX–EM algorithm parameter updates involve computing the trace of a matrix. For the new incomplete data specification this matrix is either the $b \times b$ matrix $\boldsymbol{C}^{ZZ}$ or the $b \times b$ matrix $\boldsymbol{Z}^T\boldsymbol{U}\boldsymbol{Z}\boldsymbol{C}^{ZZ}$ where $\boldsymbol{U}$ is a projection matrix. When the incomplete data is specified as $\boldsymbol{y}_o$ this matrix is either $\boldsymbol{C}^{ZZ}$ or the $n \times n$ matrix $\boldsymbol{W}\boldsymbol{C}^{-1}\boldsymbol{W}^T$ where $\boldsymbol{W} = (\boldsymbol{X} \quad \boldsymbol{Z})$. The latter involves the inverse of the entire coefficient matrix associated with Henderson's mixed model equations whereas the former only involves a smaller partition of this matrix. It is computing the trace of $\boldsymbol{W}\boldsymbol{C}^{-1}\boldsymbol{W}^T$ which makes a REML EM or REML PX–EM algorithm where the incomplete data is specified as $\boldsymbol{y}_o$ less computationally efficient than a REML EM or REML PX–EM algorithm based on the new incomplete data specification.

Although the REML EM and REML PX–EM algorithm for linear mixed models based on the new incomplete data specification is a step forward in terms of computational efficiency further work is required before either of these two algorithms can be easily implemented within existing software packages using Newton-Raphson type algorithms. In the case of the average information algorithm (Gilmour et al., 1995) the only term computed in terms of the inverse of the coefficient matrix associated with Henderson's mixed model equations is $\boldsymbol{C}^{ZZ}$. The REML EM and REML PX–EM algorithm require computing $\boldsymbol{Z}^T\boldsymbol{U}\boldsymbol{Z}\boldsymbol{C}^{ZZ}$. A potential area of future research would be to study approximations of the trace of this matrix only using terms that form part of the average information algorithm computing strategy.

The improvements to the REML EM and REML PX–EM algorithms presented in this paper make these two algorithms, particularly the latter, more likely to be implemented alongside Newton–Raphson type algorithms in statistical software packages for linear mixed models. In such a situation this would provide users of these models a viable alternative in the event of a Newton–Raphson type algorithm failing.

Acknowledgements

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

References

Callanan, T. P. and 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., and Thompson, R. (2004). Perspectives of ANOVA, REML, and a general linear mixed model. Methods and Models in Statistics, symposium held to honour the 80th birthday of Professor John Nelder FRS (Eds N. M. Adams, M. J. Crowder, D. J. Hand and D. A. Stephens), pp. 53–94. London: Imperial College.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **39,** 1–38.

Foulley, J.-L., Jafrezic, F., and 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. and 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.R., Thompson, R., and Cullis, B.R. (1995). AI, an efficient algorithm for REML estimation in linear mixed models. *Biometrics* **51,** 1440–1450.

Henderson, C.R., Kempthorne, O., Searle, S.R., and von Krosigk, C.M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15,** 192–218.

Horn, R.A., and Johnson, C.R. (1990). *Matrix Analysis.* Cambridge University Press.

Knight, E. (2008) Improved iterative schemes for REML estimation of variance parameters in linear mixed models. *Unpublished PhD Thesis.* The University of Adelaide.

Laird, N.M. and Ware, J.H. (1982). Random–Effects Models for Longitudinal Data. *Biometrics* **38,** 963–974.

Liu, C., Rubin, D., and Wu, Y. (1998). Parameter expansion to accelerate EM: The PX–EM algorithm. *Biometrika* **85,** 755–770.

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

McLachlan, G.J., and Krishnan, K. (1997). *The EM Algorithm and Extensions.* John Wiley and Sons Inc.

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.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58,** 545–554.

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