The PX-EM algorithm for fast stable fitting of Henderson's mixed model

Jean-Louis Foulley, David van Dyk

Genet. Sel. Evol. 32 (2000) 143–163

Original article

The PX-EM algorithm for fast stable fitting of Henderson's mixed model

Jean-Louis FOULLEY[a*], David A. VAN DYK[b]

[a] Station de génétique quantitative et appliquée
Institut national de la recherche agronomique
78352 Jouy-en-Josas Cedex, France
[b] Department of Statistics, Harvard University
Cambridge, MA 02138, USA

(Received 7 September 1999; accepted 3 January 2000)

Abstract – This paper presents procedures for implementing the PX-EM algorithm of Liu, Rubin and Wu to compute REML estimates of variance covariance components in Henderson's linear mixed models. The class of models considered encompasses several correlated random factors having the same vector length e.g., as in random regression models for longitudinal data analysis and in sire-maternal grandsire models for genetic evaluation. Numerical examples are presented to illustrate the procedures. Much better results in terms of convergence characteristics (number of iterations and time required for convergence) are obtained for PX-EM relative to the basic EM algorithm in the random regression.

EM algorithm / REML / mixed models / random regression / variance components

Résumé – L'algorithme PX-EM dans le contexte de la méthodologie du modèle mixte d'Henderson. Cet article présente des procédés permettant de mettre en œuvre l'algorithme PX-EM de Liu, Rubin et Wu à des modèles linéaires mixtes d'Henderson. La classe de modèles considérée concerne plusieurs facteurs aléatoires corrélés ayant la même dimension vectorielle comme c'est le cas avec les modèles de régression aléatoire dans l'analyse des données longitudinales ou avec les modèles père-grand-père maternel en évaluation génétique. Des exemples numériques sont présentés pour illustrer ces techniques. L'algorithme PX-EM présente de nettement meilleurs résultats en terme de caractéristiques de convergence (nombre d'itérations et temps de calcul) que l'EM de base sur les exemples ayant trait à des modèles de régression aléatoire.

algorithme EM / REML / modèles mixtes / régression aléatoire / composantes de variance

* Correspondence and reprints
E-mail: foulley@jouy.inra.fr

1. INTRODUCTION

Since the landmark paper of Dempster et al. [4], the EM algorithm has been among the most popular statistical techniques for calculating parameter estimates via maximum likelihood, especially in models accounting for missing data, or in models that can be formulated as such. As explained by Meng and van Dyk [23], the popularity of EM stems mainly from its computational simplicity, its numerical stability, and its broad range of applications.

Biometricians, especially those working in animal breeding, have been among the largest users of EM. Modern genetic evaluation typically relies on best linear unbiased prediction (BLUP) of breeding values [13,14] and on restricted (or residual) maximum likelihood (REML) of variance components of Gaussian linear mixed models [12, 27]. BLUP estimates are obtained by solving Henderson's mixed model equations, the elements of which are natural components of the E-step of the EM algorithm for REML estimation which explains the popularity of the triple of BLUP-EM-REML.

Unfortunately, the EM algorithm can be very slow to converge in this setting and various alternative procedures have been proposed: see e.g., Misztal's [26] review of the properties of various algorithms for variance component estimation and Johnson and Thompson [15] and Meyer [25] for a discussion of a second order algorithm based on the average of the observed and expected information.

Despite its slow convergence, EM has remained popular, primarily because of its simplicity and stability relative to alternatives [32]. Thus, much work has focused on speeding up EM while maintaining these advantages. Rescaling the random effects which are treated as missing data by EM has been a very successful strategy employed by several authors; e.g., Anderson and Aitkin [2] for binary response analysis, Foulley and Quaas [7] for heteroskedastic mixed models, and Meng and van Dyk [24] for mixed effects models using a Cholesky decomposition (see also procedures developed by Lindstrom and Bates [17] for repeated measure analysis and Wolfinger and Tobias [35] for a mixed model approach to the analysis of robust-designed experiments). To further improve computational efficiency, the principle underlying the random effects was generalized by Liu et al. [21] who introduced the parameter expanded EM or PX-EM algorithm, which in the case of mixed effects models fits the rescaling factor in the iteration.

The purpose of this paper is twofold: (i) to give an overview of this new algorithm to the biometric community, and (ii) to illustrate the procedure with several small numerical examples demonstrating the computational gain of PX-EM.

The paper is organized as follows into six sections. In Section 2, the general structure of the models is described and in Section 3 a typical EM implementation for these models (called EM_0 for clarity) is reviewed. The fourth Section briefly introduces the general PX-EM algorithm and gives appropriate formulae for mixed linear models. Two examples (sire-maternal grandsire models and random coefficient models) appear in Section 5 and Section 6 contains a brief discussion.

2. MODEL STRUCTURE

We consider the class of linear mixed models including K dependent $(u_k; k = 1, 2, \ldots, K)$ random factors, $\mathbf{u}_k$ for $k = 1, 2, \ldots, K$ using Henderson's notation,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{u}_k + \mathbf{e} \tag{1}$$

or, in a more compact form

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

where $\mathbf{y}$ is the $(N \times 1)$ data vector, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of fixed effects with matrix $\mathbf{X}$ of explanatory discrete or continuous variables, $\mathbf{u}$ is a $(q_+ \times 1)$ vector of random effects $(q_+ = \sum_{k=1}^{K} q_k)$ formed by concatenating the $K(q_k \times 1)$ vectors $\mathbf{u}_k$, $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_k', \ldots, \mathbf{u}_K')'$ with corresponding incidence matrix $\mathbf{Z}_{(N \times q_+)} = (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_k, \ldots, \mathbf{Z}_K)$, and $\mathbf{e}$ is a $(N \times 1)$ vector of residuals.

The usual Gaussian assumption is made for the distribution of $(\mathbf{y}', \mathbf{u}', \mathbf{e}')'$ i.e., $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})$ where

$$\mathbf{G} = \operatorname{var}(\mathbf{u}) = \{\mathbf{G}_{k,l}\} \quad \text{with} \quad \mathbf{G}_{k,l} = \operatorname{Cov}(\mathbf{u}_k, \mathbf{u}_l') = \mathbf{A}_{kl} g_{kl} \tag{2a}$$

and

$$\mathbf{R} = \operatorname{var}(\mathbf{e}) = \mathbf{H}\sigma_e^2 \tag{2b}$$

In (2a), $\mathbf{A}_{kl}$ is a $(q_k \times q_l)$ matrix of known coefficients and g_{kl} is a real parameter known as the (k, l) covariance component such that $\mathbf{G}_0 = \{g_{kl}\}$, the u-covariance matrix, is positive definite; a similar definition applies to $\mathbf{H}$ and σ_e^2 for the residual variance component.

We assume that all u-components have the same dimension i.e., the same number of experimental units, $q_k = q$ for all k, and similarly $\mathbf{A}_{kl} = \mathbf{A}$ for all k, l, so that $\mathbf{G}$ can be written as:

$$\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}, \tag{3}$$

where $\otimes$ symbolizes the direct or Kronecker product as defined e.g., in Searle [31].

Two important models in genetics and biometrics belong to this class of models.

First, the "sire-maternal grandsire" model (or SMGS model) as described by Bertrand and Benyshek [3],

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s \mathbf{u}_s + \mathbf{Z}_t \mathbf{u}_t + \mathbf{e}, \tag{4}$$

where $\mathbf{u}_s$ and $\mathbf{u}_t$ refer to $(q \times 1)$ vectors of sire and maternal grandsire contributions of q males respectively and $\mathbf{A}$ is the matrix of additive genetic relationships (or twice Malecot's kinship coefficients) between those males.

Here $\mathbf{R} = \mathbf{I}\sigma_e^2$, $\mathbf{G} = \begin{bmatrix} \mathbf{A}\sigma_s^2 & \mathbf{A}\sigma_{st} \\ \mathbf{A}\sigma_{st} & \mathbf{A}\sigma_t^2 \end{bmatrix} = \mathbf{G}_0 \otimes \mathbf{A}$, $\mathbf{G}_0 = \begin{bmatrix} \sigma_s^2 & \sigma_{st} \\ \sigma_{st} & \sigma_t^2 \end{bmatrix}$, and $\mathbf{g}_0 = \text{vech}\,(\mathbf{G}_0) = (\sigma_s^2, \sigma_{st}, \sigma_t^2)'$ is a function of the variance covariance components of additive direct and maternal effects of genes.

The second model, a random coefficient model [22], can be applied for instance to longitudinal data analysis (Diggle et al. [5]; Laird and Ware [16]; Schaeffer and Dekkers [30]), and is usually written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{k=1}^{K} \mathbf{Z}_{ik}\mathbf{u}_{ik} + \mathbf{e}_i \quad \text{for} \quad i = 1, 2, \ldots, q$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ij}, \ldots, y_{in_i})'$ is the $(n_i \times 1)$ vector of measurements made on the ith individual $(i = 1, 2, \ldots, q)$, $\mathbf{X}_i\boldsymbol{\beta}$ is the contribution of fixed effects, and is the kth random regression coefficient (e.g., intercept, linear slope) on covariate information $\mathbf{Z}_{ik}$ (e.g., time or age) pertaining to the ith individual.

Under the general form (1), individuals are nested within random effects whereas in random coefficient models, the opposite holds, coefficients (factors) are nested within individuals. That is,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \quad \text{for} \quad i = 1, 2, \ldots, q \tag{5}$$

with $\mathbf{u}_i = (\mathbf{u}_{i1}, \mathbf{u}_{i2}, \ldots, \mathbf{u}_{ik}, \ldots, \mathbf{u}_{iK})'$, and $\mathbf{Z}_{i(N\times K)} = (\mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \ldots, \mathbf{Z}_{ik}, \ldots, \mathbf{Z}_{iK})$ so that under (5), $\text{var}(\mathbf{u}_i) = a_{ii}\mathbf{G}_0$ and $\text{cov}\,(\mathbf{u}_i, \mathbf{u}_{i'}) = a_{ii'}\mathbf{G}_0$, i.e.,

$$\text{var}\,(\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_i', \ldots, \mathbf{u}_q')' = \mathbf{A} \otimes \mathbf{G}_0.$$

Generally, these models assume independence among residuals $\text{var}(\mathbf{e}_i) = \mathbf{I}_{n_i}\sigma_e^2$ and independence among individuals, $\mathbf{A} = \mathbf{I}_q$, but this is neither mandatory nor always appropriate as e.g., with data recorded on relatives.

Readers may be more familiar with one of the two forms (1) or (5), but both are obviously equivalent and we may use whichever is more convenient.

3. THE EM$_0$ ALGORITHM

3.1. Typical procedure

To define an EM algorithm to compute REML estimates of the model parameter $\boldsymbol{\gamma} = (\mathbf{g}_0', \sigma_e^2)'$, we hypothesize a complete data set, $\mathbf{x}$, which augments the observed data, $\mathbf{y}$, i.e. $\mathbf{x} = (\mathbf{y}', \boldsymbol{\beta}', \mathbf{u}')'$. As in [4] and [7], we treat $\boldsymbol{\beta}$ as a vector of random effects with variance tending to infinity.

Each iteration of the EM algorithm consists of two steps, the expectation or E step and the maximization or M step. In the Gaussian mixed model, this separates the computation into two simple pieces. The E-step consists of taking the expectation of the complete data log likelihood $L(\boldsymbol{\gamma}; \mathbf{x}) = \ln p(\mathbf{x}|\boldsymbol{\gamma})$ with respect to the conditional distribution of the "missing data": $\mathbf{z} = (\boldsymbol{\beta}', \mathbf{u}')'$ vector given the observed data $\mathbf{y}$ with $\boldsymbol{\gamma}$ set at its current value $\boldsymbol{\gamma}^{[t]}$ i.e.,

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = \int L(\boldsymbol{\gamma}; \mathbf{y}, \mathbf{z}) p(\mathbf{z}|\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}) \mathrm{d}\mathbf{z}, \tag{6}$$

while the M-step updates $\boldsymbol{\gamma}$ by maximizing (6) with respect to $\boldsymbol{\gamma}$ i.e.,

$$\boldsymbol{\gamma}^{[t+1]} = \arg\max_{\boldsymbol{\gamma}} Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}). \tag{7}$$

We begin by deriving an explicit expression for (6) and then derive the two steps of the EM algorithm. By definition $p(\mathbf{x}|\boldsymbol{\gamma}) = p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\gamma})p(\boldsymbol{\beta}, \mathbf{u}|\boldsymbol{\gamma})$, where

$$p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\gamma}) = p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2) = p(\mathbf{e}|\sigma_e^2)$$

and

$$p(\boldsymbol{\beta}, \mathbf{u}|\boldsymbol{\gamma}) \propto p(\mathbf{u}|\mathbf{g}_0),$$

so that

$$L(\boldsymbol{\gamma}; \mathbf{x}) = L(\sigma_e^2; \mathbf{e}) + L(\mathbf{g}_0; \mathbf{u}) + const. \tag{8}$$

Formula (8) allows the formal dissociation of the computations pertaining to the residual σ_e^2 from the u-components of variance $\mathbf{g}_0$. Combining (6) with (8) leads to

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = Q_e(\sigma_e^2|\boldsymbol{\gamma}^{[t]}) + Q_u(\mathbf{g}_0|\boldsymbol{\gamma}^{[t]}) + const. \tag{9}$$

The expressions on the right-hand side can be written explicity

$$Q_e(\sigma_e^2|\boldsymbol{\gamma}^{[t]}) = -1/2[N\ln 2\pi + \ln|\mathbf{H}| + N\ln\sigma_e^2 + E(\mathbf{e}'\mathbf{H}^{-1}\mathbf{e}|\mathbf{y}, \boldsymbol{\gamma}^{[t]})/\sigma_e^2] \tag{10}$$

and,

$$Q_u(\mathbf{g}_0|\boldsymbol{\gamma}^{[t]} = -1/2[q_+\ln 2\,\pi + \ln|\mathbf{G}| + E(\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}|\mathbf{y}, \boldsymbol{\gamma}^{[t]})].$$

Under assumption (3), $|\mathbf{G}| = |\mathbf{G}_0|^q|\mathbf{A}|^K$ and $\mathbf{G}^{-1} = \mathbf{G}_0^{-1} \otimes \mathbf{A}^1$, so $Q_u(\mathbf{g}_0|\boldsymbol{\gamma}^{[t]})$ reduces to

$$Q_u(\mathbf{g}_0|\boldsymbol{\gamma}^{[t]}) = -1/2[q_+\ln 2\pi + K\ln|\mathbf{A}| + q\ln|\mathbf{G}_0| + \mathrm{tr}\,(\mathbf{G}_0^{-1}\Omega^{[t]})], \tag{11}$$

where $\boldsymbol{\Omega}^{[t]} = E\{\mathbf{u}_k'\mathbf{A}^{-1}\mathbf{u}_l|\mathbf{y}, \boldsymbol{\gamma}^{[t]}\}$ for $k, l = 1, 2, \ldots, K$.

For the M-step, we maximize (10) as a function of σ_e^2, and (11) as a function of $\mathbf{g}_0$. For $\mathbf{H}$ known, this results in

$$\sigma_e^{2[t+1]} = [E(\mathbf{e}'\mathbf{H}^{-1}\mathbf{e}|\mathbf{y}, \boldsymbol{\gamma}^{[t]})]/N \tag{12}$$

and

$$\mathbf{G}_0^{[t+1]} = \Omega^{[t]}/q, \tag{13}$$

see Lemma 3.2.2 of Anderson [1], page 62.

The expectations in (12) and (13), i.e. the E-step, can be computed using elements of Henderson's [14] mixed model equations (ignoring subscripts), i.e.,

$$\begin{bmatrix} \mathbf{X}'\mathbf{H}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{H}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{H}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{H}^{-1}\mathbf{Z} + \sigma_e^2\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{H}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{H}^{-1}\mathbf{y} \end{bmatrix}, \tag{14}$$

where $\hat{\boldsymbol{\beta}}$ is the GLS estimate of $\boldsymbol{\beta}$, and $\hat{\mathbf{u}}$ is the BLUP of $\mathbf{u}$. In particular, we compute

$$E(\mathbf{e}'\mathbf{H}^{-1}\mathbf{e}|\mathbf{y},\boldsymbol{\gamma}) = \hat{\mathbf{e}}'\mathbf{H}^{-1}\hat{\mathbf{e}} + \sigma_e^2[p + q_+ - \sigma_e^2 \mathrm{tr}\,(\mathbf{C}_{uu}\mathbf{G}^{-1})], \tag{15}$$

and

$$E(\mathbf{u}_k'\mathbf{A}^{-1}\mathbf{u}_l|\mathbf{y},\boldsymbol{\gamma}) = \hat{\mathbf{u}}_k'\mathbf{A}^{-1}\hat{\mathbf{u}}_l + \sigma_e^2\,\mathrm{tr}\,(\mathbf{A}^{-1}\mathbf{C}_{u_k u_l}), \tag{16}$$

where $p = \mathrm{rank}(\mathbf{X})$, $q_+ = Kq = \dim(\mathbf{u})$ and $\mathbf{C}_{uu}$ is the block of the inverse of the coefficient matrix of (14) corresponding to $\mathbf{u}$. Further numerical simplifications can be carried out to avoid inverting the coefficient matrix at each iteration using diagonalization or tridiagonalization procedures (see e.g., Quaas [29]).

3.2. An ECME version

In order to improve computational performance, we can sometimes update some parameters without defining a complete data set. In particular, the ECME algorithm [20] suggests separating the parameter into several sub parameters (i.e. , model reduction), and updating each sub parameter in turn conditional on the others. For each of these sub parameters, we can maximize either the observed data log likelihood directly, i.e., $L(\boldsymbol{\gamma};\mathbf{y})$ or the expected augmented data log likelihood, $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]})$.

To implement an ECME algorithm in the mixed effects model, we rewrite the parameter as $\boldsymbol{\zeta} = (\mathbf{d}_0', \sigma_e^2)$ where $\mathrm{var}(\mathbf{y}) = \mathbf{W}\sigma_e^2$ with $\mathbf{W} = \mathbf{ZDZ}' + \mathbf{H}$, $\mathbf{D} = \mathbf{D}_0 \otimes \mathbf{A}$, and $\mathbf{d}_0 = \mathrm{vech}(\mathbf{D}_0)$ and first update σ_e^2 by directly maximizing $L(\boldsymbol{\zeta};\mathbf{y})$ (without recourse to missing data) under the constraint that $\mathbf{d}_0$ is fixed at $\mathbf{d}_0^{[t]}$

$$\sigma_e^{2[t+1]} = \frac{[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\mathbf{d}_0^{[t]})]'[\mathbf{W}(\mathbf{d}_0^{[t]})]^{-1}[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\mathbf{d}_0^{[t]})]}{N - p}. \tag{17}$$

An ML analogue of this formula (dividing by N instead of $N - p$) was first obtained by Hartley and Rao [12] in their general ML estimation approach to the parameters of mixed linear models; see also Diggle et al [5], page 67. Second, we update $\mathbf{d}_0$ by maximizing $Q(\mathbf{d}_0, \sigma^{2[t+1]}|\boldsymbol{\zeta}^{[t]})$ using the missing data approach,

$$\mathbf{D}_0^{[t+1]} = \mathbf{G}_0^{[t+1]}/\sigma_e^{2[t+1]}, \tag{18}$$

where $\mathbf{G}_0^{[t+1]}$ is defined in (13).

Henderson [13] showed that the expression for σ_e^2 in (17) can be obtained from his mixed model equations solutions as

$$\sigma_e^{2[t+1]} = \frac{[\mathbf{y}'\mathbf{H}^{-1}\mathbf{y} - \hat{\boldsymbol{\beta}}^{[t]'}\mathbf{X}'\mathbf{H}^{-1}\mathbf{y} - \hat{\mathbf{u}}^{[t]'}\mathbf{Z}'\mathbf{H}^{-1}\mathbf{y}]}{N - p}, \tag{19}$$

where $\hat{\boldsymbol{\beta}}^{[t]}$ and $\hat{\mathbf{u}}^{[t]}$ are defined by (14) evaluated with $\sigma_e^2\mathbf{G}^{-1} = \mathbf{D}^{-1}$ computed using $\mathbf{d}_0^{[t]}$. Incidentally, this shows that the algorithm developed by Henderson [14] to compute REML estimates, as early as 1973, introduces model reduction in a manner similar to recent EM-type algorithms.

4. THE PX-EM ALGORITHM

4.1. Generalities

In the PX-EM algorithm proposed by Liu et al. [21], the parameter space of the complete data model is expanded to a larger set of parameters, $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_*, \boldsymbol{\alpha})$, with $\boldsymbol{\alpha}$ a working parameter, such that $(\boldsymbol{\gamma}_*, \boldsymbol{\alpha})$ satisfies the following two conditions:

– it can be reduced to the original parameter $\boldsymbol{\gamma}$, maintaining the observed data model via a many-to-one reduction form $\boldsymbol{\gamma} = R(\boldsymbol{\Gamma})$;

– when $\boldsymbol{\alpha}$ is set to its reference (or "null") value, $(\boldsymbol{\gamma}_*, \boldsymbol{\alpha}_0)$ induces the same complete data model as with $\boldsymbol{\gamma} = \boldsymbol{\gamma}_*$ i.e., $p[\mathbf{x}|\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_*, \boldsymbol{\alpha}_0)] = p[\mathbf{x}|\boldsymbol{\gamma} = \boldsymbol{\gamma}_*]$.

We introduce the working parameter because the original EM (EM_0) imputes missing data under a wrong model, i.e., the EM iterate $\boldsymbol{\gamma}_{EM}^{[t]}$ is different from the MLE. The PX algorithm takes advantage of the difference between the imputed value $\boldsymbol{\alpha}^{[t+1]}$ of $\boldsymbol{\alpha}$ and its reference value $\boldsymbol{\alpha}_0$ to make what Liu et al. [21] called a covariance adjustment in $\boldsymbol{\gamma}$, i.e.,

$$\boldsymbol{\gamma}_X^{[t+1]} - \boldsymbol{\gamma}_{EM}^{[t+1]} \approx \mathbf{b}_{\gamma|\alpha}(\boldsymbol{\alpha}^{[t+1]} - \alpha_0) \tag{20}$$

where $\boldsymbol{\gamma}_X^{[t]}$ is the PX-EM value at iteration $[t]$, $\boldsymbol{\gamma}_{EM}^{[t]}$ is the EM iterate, and $\mathbf{b}_{\gamma|\alpha}$ is a correction factor. Liu et al. [21] show that this adjustment necessarily improves the rate of convergence of EM generally in terms of the number of iterations required for convergence.

Operationally, the PX-EM algorithm, like EM, consists of two steps. In particular, the PX-E step computes the conditional expectation of the log likelihood of $\mathbf{x}$ given the observed data $\mathbf{y}$ with $\boldsymbol{\Gamma}^{[t]}$ set to $(\boldsymbol{\gamma}_*^{[t]}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_0)$ i.e.,

$$Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{[t]}) = E[L(\boldsymbol{\Gamma}; \mathbf{x})|\mathbf{y}, \boldsymbol{\Gamma}^{[t]} = (\boldsymbol{\gamma}_*^{[t]}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_0)]. \tag{21}$$

The PX-M step then maximizes (21) with respect to the expanded parameters

$$\boldsymbol{\Gamma}^{[t+1]} = \arg\max_{\Gamma} Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{[t]}), \tag{22}$$

and $\boldsymbol{\gamma}$ is updated via $\boldsymbol{\gamma}^{[t+1]} = R(\boldsymbol{\Gamma}^{[t+1]})$.

In the next section, we illustrate PX-EM in the Gaussian linear model. In particular, we will describe a simple method of introducing a working parameter into the complete data model.

4.2. Implementation of PX-EM in the mixed model

We begin by defining the working parameter as a $(K \times K)$ invertible real matrix $\boldsymbol{\alpha} = \{\alpha_{kl}\}$ which we incorporate into the model by rescaling the random effects $\tilde{\mathbf{U}} = \boldsymbol{\alpha}^{-1}\mathbf{U}$ where $\mathbf{U}_{(K\times q)} = (\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_k, \ldots, \mathbf{u}_K)'$ i.e.,

$$\begin{cases} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{k=1}^{K} \mathbf{Z}_k \mathbf{u}_k + \mathbf{e} \\ \mathbf{u}_k = \sum_{l=1}^{K} \alpha_{kl} \tilde{\mathbf{u}}_l \end{cases} \tag{23a}$$

or alternatively, under (5)

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\alpha}\tilde{\mathbf{u}}_i + \mathbf{e}_i. \tag{23b}$$

By the definition of $\tilde{\mathbf{u}}_i$, we have $\tilde{\mathbf{u}}_i \sim N(\mathbf{0}, \mathbf{G}_{0^*})$, where $\mathbf{G}_{0^*} = \boldsymbol{\alpha}^{-1}\mathbf{G}_0(\boldsymbol{\alpha}^{-1})'$. Rescaling the random effects by $\boldsymbol{\alpha}$ introduces the working parameters into two parts of the model, into (23) and into the distribution of $\tilde{\mathbf{u}}_i$ which can be viewed as an extended parametric form of the distribution of $\mathbf{u}_i$ (see (2a)) i.e., when $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0 = \mathbf{I}_K$, $\mathbf{u}_i$ and $\tilde{\mathbf{u}}_i$ have the same distribution.

To understand why the PX-EM works in this case, recall that the REML estimate is the value of which maximizes

$$L(\boldsymbol{\gamma}; \mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\gamma}) p(\mathbf{u}|\boldsymbol{\gamma}) \mathrm{d}\boldsymbol{\beta}\mathrm{d}\mathbf{u}. \tag{24a}$$

What is important here is that for any value of $\boldsymbol{\alpha}$, $L(\boldsymbol{\gamma}; \mathbf{y}) = L(\boldsymbol{\gamma}_*, \boldsymbol{\alpha}; \mathbf{y})$ with

$$L(\boldsymbol{\gamma}_*, \boldsymbol{\alpha}; \mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\beta}, \tilde{\mathbf{u}}, \boldsymbol{\gamma}_*, \boldsymbol{\alpha}) p(\tilde{\mathbf{u}}|\boldsymbol{\gamma}_*) \mathrm{d}\boldsymbol{\beta}\mathrm{d}\tilde{\mathbf{u}}. \tag{24b}$$

Thus, we can fix $\boldsymbol{\alpha}$ in (24b) to be any value at each iteration. Liu et al. [21] showed that fitting $\boldsymbol{\alpha}$ in the iteration can improve the computational performance of the algorithm.

Computationally, the PX-EM algorithm replaces the integrand of (24a) with that of (24b). In particular, in the E-step, we compute $Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{[t]}) = (\boldsymbol{\gamma}_*^{[t]}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_0)$. Here we choose $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, since any value of $\boldsymbol{\alpha}$ will work and using $\boldsymbol{\alpha}_0$ reduces computations to those of the EM_0 algorithm. In the M-step, we update $\boldsymbol{\Gamma}$ by maximizing $Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{[t]})$, i.e. we compute $\mathbf{g}_{0^*}^{[t+1]} = \mathrm{vec}(\mathbf{G}_{0^*}^{[t+1]})$, $\boldsymbol{\alpha}^{[t+1]}$ and $\sigma_{e^*}^{2[t+1]}$. Finally, we reduce these parameter values to those of interest, $\mathbf{G}_0^{[t+1]} = \boldsymbol{\alpha}^{[t+1]}\mathbf{G}_{0^*}^{[t+1]}(\boldsymbol{\alpha}^{[t+1]})'$ and $\sigma_e^{2[t+1]} = \sigma_{e^*}^{2[t+1]}$ (in the remainder of the paper we fix $\sigma_{e^*}^2$ at σ_e^2). We now move to the details of these computations.

In order to derive $Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{[t]}$, we note that

$$p(\mathbf{y}, \boldsymbol{\beta}, \tilde{\mathbf{u}}|\mathbf{g}_{0^*}, \boldsymbol{\alpha}, \sigma_e^2) \propto p(\mathbf{y}|\boldsymbol{\beta}, \tilde{\mathbf{u}}, \boldsymbol{\alpha}, \sigma_e^2) p(\tilde{\mathbf{u}}|\mathbf{g}_{0^*}),$$

and

$$L(\boldsymbol{\Gamma}|\mathbf{x}) = L(\boldsymbol{\alpha}, \sigma_e^2|\mathbf{e}) + L(\mathbf{g}_{0^*}|\tilde{\mathbf{u}}) + const. \tag{25}$$

thus

$$Q(\boldsymbol{\Gamma}|\boldsymbol{\Gamma}^{[t]}) = Q_e(\boldsymbol{\alpha}, \sigma_e^2|\boldsymbol{\Gamma}^{[t]}) + Q_u(\mathbf{g}_{0^*}|\boldsymbol{\Gamma}^{[t]}) + const. \tag{26}$$

where $\boldsymbol{\Gamma} = (\mathbf{g}_{0^*}', (\mathrm{vec}\,\boldsymbol{\alpha})', \sigma_e^2)'$, and $\boldsymbol{\Gamma}^{[t]} = (\mathbf{g}_{0^*}^{[t]'}, (\mathrm{vec}\,\boldsymbol{\alpha}_0)', \sigma_e^{2[t]})'$.
Maximizing $Q_u(\mathbf{g}_{0^*}|\boldsymbol{\Gamma}^{[t]})$ with respect to $\mathbf{g}_{0^*}$ is identical to the corresponding calculations for $\mathbf{g}_0$ in EM_0 i.e., we set $\mathbf{G}_{0^*}^{[t+1]} = \boldsymbol{\Omega}^{[t]}/q$, where $\boldsymbol{\Omega}^{[t]}$ is evaluated as in EM_0 with (16).

Next, we wish to maximize $Q_e(\boldsymbol{\alpha}, \sigma^2_{e^*}|\boldsymbol{\Gamma}^{[t]})$, which can be written formally as in (10) but with $\mathbf{e}$ defined in (23a) or (23b). Partial derivatives of this function with respect to $\boldsymbol{\alpha}$ are given by:

$$\frac{\partial Q}{\partial \alpha_{kl}} = \frac{1}{\sigma^2_e} \sum_{i=1}^{I} E\left[\mathbf{u}'_i \frac{\partial \boldsymbol{\alpha}}{\partial \alpha_{kl}} \mathbf{Z}'_i \mathbf{H}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \boldsymbol{\alpha} \tilde{\mathbf{u}}_i) | \mathbf{y}, \boldsymbol{\Gamma}^{[t]} \right].$$

Solving these K^2 equations does not involve σ^2_e, and is equivalent to solving the linear system $\mathbf{F}(\operatorname{vec} \boldsymbol{\alpha}') = \mathbf{h}$,

$$\sum_{m=1}^{K} \sum_{n=1}^{K} f^{[t]}_{kl,mn} \alpha^{[t+1]}_{mn} = h^{[t]}_{kl}; \text{ for } \; k, l = 1, 2, \ldots, K \tag{27}$$

where

$$f^{[t]}_{kl,mn,} = \operatorname{tr}\left[\mathbf{Z}'_k \mathbf{H}^{-1} \mathbf{Z}_m E(\mathbf{u}_n \mathbf{u}'_l | \mathbf{y}, \boldsymbol{\Gamma}^{[t]}) \right]] \tag{28}$$

and

$$h^{[t]}_{kl} = \operatorname{tr}\left\{ \mathbf{Z}'_k \mathbf{H}^{-1} E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \mathbf{u}'_l | \mathbf{y}, \boldsymbol{\Gamma}^{[t]}] \right\}. \tag{29}$$

Explicit expressions for the coefficients when $K = 2$ are given in Appendix A.

We can compute $\boldsymbol{\alpha}^{[t+1]}$ using Henderson's [14] mixed model equations (14), suppressing the superscript $[t]$, as follows

$$f_{kl,mn} = \operatorname{tr}\left[\mathbf{Z}'_k \mathbf{H}^{-1} \mathbf{Z}_m (\hat{\mathbf{u}}_n \hat{\mathbf{u}}'_l + \sigma^2_e \mathbf{C}_{u_n u_l}) \right], \tag{30}$$

$$h_{kl} = \hat{\mathbf{u}}'_l \mathbf{Z}'_k \mathbf{H}^{-1} \mathbf{y} - \operatorname{tr}\left[\mathbf{Z}'_k \mathbf{H}^{-1} \mathbf{X} (\hat{\boldsymbol{\beta}} \hat{\mathbf{u}}'_l + \sigma^2_e \mathbf{C}_{\beta u_l}) \right], \tag{31}$$

where $\mathbf{Z}'_k \mathbf{H}^{-1} \mathbf{Z}_m$ is the block of the coefficient matrix corresponding to $\mathbf{u}_k$ and $\mathbf{u}_m$; $\mathbf{Z}'_k \mathbf{H}^{-1} \mathbf{X}$ is the block corresponding to $\mathbf{u}_k$ and $\boldsymbol{\beta}$; $\mathbf{C}_{u_k u_m}$ and $\mathbf{C}_{u_k \beta} = \mathbf{C}'_{\beta u_k}$ are the corresponding blocks in the inverse coefficient matrix; $\mathbf{Z}'_k \mathbf{H}^{-1} \mathbf{y}$ is the sub vector in the right hand side of (14) corresponding to $\mathbf{u}_k$; and $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}_k$ are the solutions for $\boldsymbol{\beta}$ and $\mathbf{u}_k$ in (14). Once we obtain $\boldsymbol{\alpha}^{[t+1]}$, we update $\mathbf{G}_0$ as indicated previously.

Finally to update σ^2_e, we maximize $Q_e(\boldsymbol{\alpha}, \sigma^2_e|\boldsymbol{\Gamma}^{[t]})$ via

$$\sigma^{2[t+1]}_e = E(\mathbf{e}' \mathbf{H}^{-1} \mathbf{e} | \mathbf{y}, \boldsymbol{\Gamma}^{[t]})/N, \tag{32}$$

where the residual vector, $\mathbf{e}$, is adjusted for the solution $\boldsymbol{\alpha}^{[t+1]}$ in (27), i.e., using $\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \boldsymbol{\alpha}^{[t+1]} \tilde{\mathbf{u}}_i$. A short-cut procedure implements a conditional maximization with $\boldsymbol{\alpha}$ fixed at $\boldsymbol{\alpha}^{[t]} = \mathbf{I}_K$ and results in formula (15) as in the EM_0 procedure. One can also derive a parameter expanded ECME algorithm by applying Henderson's formula (19) with $\mathbf{d}_0$ fixed at $\mathbf{d}^{[t]}_0$; see van Dyk [32].

5. NUMERICAL EXAMPLES

5.1. Description

In this section, we illustrate the procedures and their computational advantage relative to more standard methods using two sire-maternal grandsire (model 4) and two random coefficient (model 5) examples.

5.1.1. Sire-maternal grandsire models

The two examples in this section are based on calving score of cattle [6]. From a biological viewpoint, parturition difficulty is a typical example of a trait involving direct effects of genes transmitted by parents on offspring, and maternal effects influencing the environmental conditions of the foetus during gestation and at parturition. Thus, statistically we must consider the sire and the maternal grandsire contributions of a male, not as simple multiples of each other (i.e., the first twice that of the second), but as two different but correlated variables. This model can be written as

$$y_{ijklm} = \mu + \alpha_i + \beta_j + s_k + t_l + e_{ijklm}, \tag{33}$$

where μ is an overall mean; α_i, β_j are fixed effects of the factors A = sex ($i = 1, 2$ for bull and heifer calves respectively) and B = parity of dam ($j = 1, 2$ and 3 for heifer, second and third calves respectively); s_k is the random contribution of male k as a sire and t_l that of male l as maternal grandsire, and e_{ijklm} are residual errors, assumed iid- $N(0, \sigma_e^2)$.

Letting $\mathbf{s} = \{s_k\}$ and $\mathbf{t} = \{t_l\}$, it is assumed that $\text{var}(\mathbf{s}) = \mathbf{A}\sigma_s^2$, $\text{var}(\mathbf{t}) = \mathbf{A}\sigma_t^2$ and $\text{cov}(\mathbf{s}, \mathbf{t}') = \mathbf{A}\sigma_{st}$, where $\mathbf{A}$ is the matrix of genetic relationships among the several males occurring as sires and maternal grandsires and $\mathbf{g}_0 = (\sigma_s^2, \sigma_{st}, \sigma_t^2)'$ is the vector of variance-covariance components.

We analyse two data sets; the first is the original data set presented in [6], which we refer to as "Calving Data 1 or CD1", and the second is a data set with the same design structure but with smaller subclass size and simulated data which we refer to as "Calving Data 2 or CD2" (see Append. B1)

5.1.2. Random coefficient models

Growth data: We first analyse a data set due to Pothoff and Roy [28] which contains facial growth measurements recorded at four ages (8, 10, 12 and 14 years) in 11 girls and 16 boys. There are nine missing values, which are defined in Little and Rubin [19]. The data appear in Verbeke and Molenberghs [33] (see Table 4.11, page 173 and Appendix B2) with a comprehensive statistical analysis.

We consider model 6 of Verbeke and Molenberghs which is a typical random coefficient model for longitudinal data analysis with an intercept and a linear slope, and can be written as

$$y_{ijk} = \mu + \alpha_i + \beta_i t_j + a_{ik} + b_{ik} t_j + e_{ijk}, \tag{34}$$

where the systematic component of the outcome y_{ijk} involves an intercept $\mu+\alpha_i$ varying according to sex i ($i = 1,\ 2$ for female and male children respectively) and a linear increase with time ($t_j = 8,\ 10,\ 12$ and 14 years); the rate β_i also varies with sex. The a_{ik} and b_{ik} are the random homologues of α_i and β_i but are defined at the individual level (indexed k within sex i). Letting $\mathbf{u}_{ik} = (a_{ik}, b_{ik})'$, it is assumed that the $\mathbf{u}_{ik}$'s are iid- $N(\mathbf{0}, \mathbf{G}_0)$. Similarly, letting $\mathbf{e}_{ik} = (e_{i1k}, e_{i2k}, e_{i3k}, e_{i4k})'$ the $\mathbf{e}_{ik}$ are assumed iid- $N(\mathbf{0}, \sigma_e^2\mathbf{I}_4)$ and distributed independently from the $\mathbf{u}_{ik}$'s.

Ultrafiltration data: In our second random coefficient model, we consider the ultrafiltration response of 20 membrane dialysers measured at 7 different transmembrane pressures with an evaluation made at 2 different blood flow rates. These data due to Vonesh and Carter [34] are described and analysed in detail in the SAS manual for mixed models (Littell et al. [18]; data set 8.2 DIAL Appendix 4, pages 575-577 ; Appendix B3).

Using notations similar to the model for the growth data, we write

$$y_{ijk} = \mu + \alpha_i + \sum_{r=1}^{4} \beta_r x_{ijk}^r + a_{ik} + \sum_{r=1}^{2} b_{r,ik} x_{ijk}^r + e_{ijk}, \tag{35}$$

where y_{ijk} is the ultrafiltration rate in $\mathrm{ml} \cdot \mathrm{h}^{-1}$, $\mu + \alpha_i$ the intercept for blood rate i ($i = 1,\ 2$ for 200, 300 $\mathrm{dl{\cdot}min^{-1}}$), $\sum_{r=1}^{4} \beta_r x_{ijk}^r$ is the regression of the response on the transmembrane pressure x_{ijk} (dm Hg) as a homogeneous quartic polynomial; a_{ik} and $b_{r,ik}$ represent the random coefficients up to the second degree of the regression defined at the dialyser level ($k = 1,\ 2, \ldots,\ 20$).

Again, letting $\mathbf{u}_{ik} = (a_{ik}, b_{1,ik}, b_{2,ik})'$ and $\mathbf{e}_{ik} = \{e_{ijk}\}$, it is assumed that the $\mathbf{u}_{ik}$'s are iid- $N(\mathbf{0}, \mathbf{G}_0)$ and the $\mathbf{e}_{ik}$'s are iid- $N(\mathbf{0}, \sigma_e^2\mathbf{I}_7)$ and distributed independently of each other.

5.2. Calculations and results

REML estimates of variance-covariance parameters were computed for each of these four data sets using the EM_0 and PX-EM procedures including the complete (PX-C) and the triangular (PX-T) working parameter matrices, (see van Dyk [32] for discussion of the use of a lower triangular matrix as the working parameter). For each of the three EM-type algorithms, the standard procedure described in this paper was applied as well as those based on Henderson's formula for the residual variance. The iteration stopped when the norm $\sqrt{\left(\sum_i \Delta\theta_i^2\right) / \left(\sum_i \theta_i^2\right)}$ of both $\mathbf{g}_0$ and of σ_e^2, was smaller than 10^{-8}.

Results are summarized in Tables I and II for sire-maternal grandsire (SMGS) and random coefficient (RC) models, respectively. In all cases, the number of iterations required for convergence was smaller for the PX-C procedure than for the EM_0 procedure. The decrease in this number of iterations was 15 to 20% for SMGS models and as high as 70% for RC models.

Table I. Performance of EM algorithms for estimating variance components in sire and maternal grandsire models.

Algorithms[2]		Examples[1] Calving data I −2L[3]	Iterations[4]	Time[5]	Calving data II −2L[3]	Iterations[4]	Time[5]
EM_0	a	1760.284442	122	5″	746.6163445	627	25″
	b	1760.284442	123	4″	746.6163444	628	22″
PX-C	a	1760.284442	98	12″	746.6163444	533	1'05″
	b	1760.284442	98	9″	746.6163444	534	1'03″
PX-T	a	1760.284442	104	13″	746.6163445	586	1'13″
	b	1760.284442	104	12″	746.6163444	586	1'09″

[1] Examples: Calving data I: Foulley [6]; Calving difficulty score shown in appendix B1. Calving data II: same design structure, 4 scores, small progeny group size and simulated data.

[2] Algorithms: EM_0: Model 0 and PX-C: PX EM with extended and complete parameters as defined in Liu et al. [21]. PX-T: PX EM with triangular matrix of parameters as defined by van Dyk [32].
a) standard procedure; b) using Henderson's formula for residual variance.

[3] −2L = minimum of minus twice the restricted log likelihood. Estimates of parameters are:
Large pgs: residual = 0.50790017; sire var = 0.03201508; sire mgs covar = 0.01146468; mgs va = 0.06304075.
Small pgs: residual = 0.852018; sire var = 0.704462; sire mgs covar = 0.533585; mgs va = 0.923661.

[4] Iterations up to convergence for a norm of both the residual variance and the matrix of u components of variance lower than or equal to 10^{-8}.

[5] Time to convergence based on an APL2 programme (Dyalog73) run on a PC Pentium I (90 MHz).

The absolute numbers of iterations were 64 vs. 224 for PX-C vs. EM_0 for growth data and 76 vs. 259 for the ultrafiltration data.

Since computing time per iteration is larger for PX-C than for EM_0, the total computing time remains smaller with EM_0 for the CD1 and CD2 examples. However, in the RC models, computing time is about halved in both examples. This impressive advantage of PX-C vs. EM_0 was also observed at different norm values (10^{-6}, 10^{-9}) and with a different stopping rule (decrease in -2L equal to or smaller than 10^{-8}); see also the plot of EM sequences for the intercept variance in the "growth data" example (Fig. 1) and for all the u-components of variance in the ultrafiltration data example (Fig. 2).

Table II. Performance of EM algorithms for estimating variance components in random coefficient models.

Algorithms[2]		Examples[1] Growth data (1st degree)			Ultrafiltration data (2nd degree)		
		$-2L$[3]	Iterations[4]	Time[5]	$-2L$[3]	Iterations[4]	Time[5]
EM_0	a	842.3559004	224	1'07″	645.8495069	259	2'30″
	b	842.3559008	241	1'04″	645.8495097	264	2'01″
PX-C	a	842.3559007	64	32″	645.8495076	76	1'18″
	b	842.3559007	67	31″	645.8495062	75	1'19″
PX-T	a	842.3559007	92	44″	645.8495083	352	5'57″
	b	842.3559007	96	43″	645.8495096	372	5'57"

[1] Examples: Growth data from Pothoff and Roy [28] with the 9 missing values defined by Little and Rubin [19]: see also Verbeke and Molenberghs [33] for a detailed analysis. Here 1st degree polynomial for the random part (intercept + slope). Ultrafiltration rates of 20 membrane dialyzers measured at 7 pressures and using two blood flood rates [18, 34]. Here, second degree polynomial for the random part.

[2] Algorithms: EM_0 : Model 0 and PX-C: PX EM with extended and complete parameters as defined in Liu et al. [21]. PX-T: PX EM with triangular matrix of parameters as defined by van Dyk [32].
a) standard procedure; b) using Henderson's formula for residual variance.

[3] $-2L$ = minimum of minus twice the restricted loglikelihood Estimates of parameters are:
Growth data (original values $\times$ 100): residual = 176.6555; (00) = 835.5160; (01) = -46.5266, (11) = 4.4150.
Ultrafiltration data: residual = 3.317524; (00) = 2.246091; (01) = -3.731253; (02) = 0.687083; (11) = 24.080699; (12) = -6.829680; (22) = 2.172312
0, 1, 2 stand for intercept, first and second degree random coefficient effects respectively.

[4] Iterations up to convergence for a norm of both the residual variance and the matrix of u components of variance lower or equal to 10^{-8}.

[5] Time to convergence based on an APL2 programme (Dyalog73) run on a PC Pentium I (90 Mhz).

Results obtained with PX-T were consistently poorer than with PX-C and in the case of ultrafiltration data even poorer than EM_0. Computation for this model is especially demanding because of an adjustment of a second degree polynomial with a very high absolute correlation between the first and second degree coefficients of 0.94.

Finally, no practical differences were observed between the standard and the ECME procedures; Henderson's formula can be used in practice to compute the residual variance.

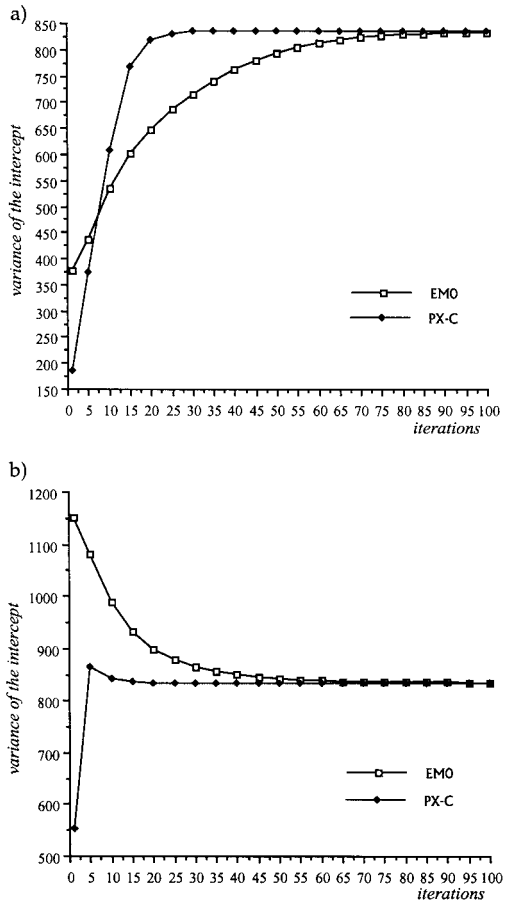

Figure 1. Two typical sequences of EM iterates for the "growth data" example.
a) starting values: residual = total variance; (00) = 500, (01) = 0, (11) = 5.
b) starting values: residual = 0.5 total variance; (00) = 2000, (01) = 0, (11) = 20.

6. DISCUSSION-CONCLUSION

This paper shows that the PX-EM procedure can be easily implemented within the framework of Henderson's mixed model equations. Changes to the EM_0 procedure are simple requiring only the solution of a $(K^2 \times K^2)$ linear system at each iteration with K generally 2, 3 or 4. In particular, the PX-EM procedure can handle the situation of random effects correlated among experimental units (e.g., individuals) as it often occurs in genetics.

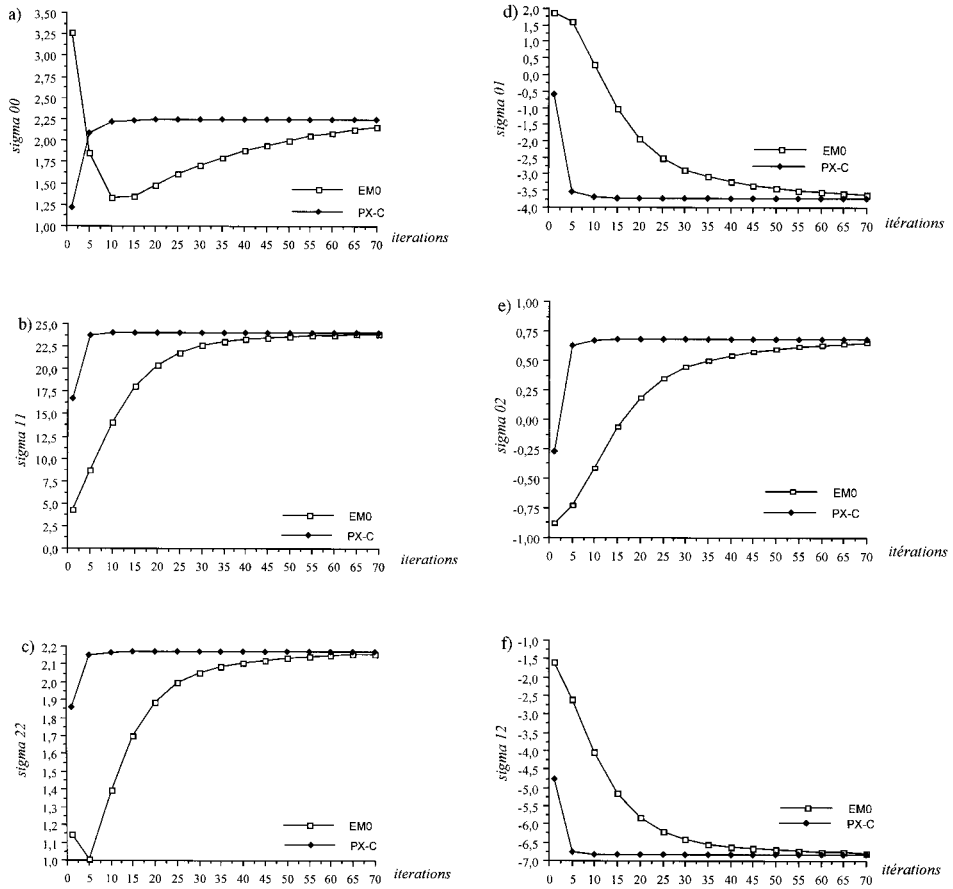

Figure 2. EM iterates for the "ultrafiltration data" example.

a, b, c, d, e and f stand for (00), (11), (22), (01), (02) and (12) u-components of variance and covariance respectively; starting values: residual = 4; (00) = (11) = (22) = 4, (01) = 2, (02) = −1.2, (12) = −2.4.

A ML extension of the PX-EM procedure is straightforward with an appropriate change in the conditional expectations of $\mathbf{u}_k\mathbf{u}_l'$ and $\boldsymbol{\beta}\mathbf{u}_k'$ along the lines proposed by Foulley et al. [8] and van Dyk [32].

Our examples confirm the potential advantage of the PX-EM procedure for mixed linear models already advocated by Liu et al. [21] and van Dyk [32]. The improvement was especially decisive in the case of random coefficient models. This is important in practice since these models are becoming more and more popular among biometricians e.g., epidemiologists and geneticists involved in the analysis of longitudinal data and space-time phenomena.

In this manuscript, emphasis was on EM procedures and ways to improve them. This does not preclude using alternative algorithms for computing maximum likelihood estimations of variance component e.g. the Average Information (AI)-REML (Gilmour, Thompson and Cullis [10]).

Nonetheless, EM procedures have two important features:

(i) they allow separation of the computations required for the **R** matrix (errors but also time or space processes) and the **G** matrix for which the PX version turns out to perform especially well; here we suppose the elements of **H** in **R** fixed but the EM procedure can be easily extended to parameters in an unknown **H** [9];

(ii) (PX)EM generally selects the correct sub model when estimates are on or near the boundary of the parameter space, a property that second order algorithms do not always exhibit (e.g., the analysis of the growth data in Foulley et al [9] and the simulations in Meng and van Dyk [24] and van Dyk, [32]).

Actually, this is an example of the well-known stable convergence of EM type algorithms (i.e., monotone convergence) which second order algorithms do not generally exhibit [32].

ACKNOWLEDGEMENTS

David van Dyk gratefully acknowledges funding for this project partially provided by the National Science Foundation (USA) grant DMS-97-05157 and the US Bureau of the Census. Thanks are also expressed to Christèle Robert-Granié and Barbara Heude for their help in the numerical validation of the growth and ultrafiltration examples via SAS proc mixed.

REFERENCES

[1] Anderson T.W., An introduction to multivariate statistical analysis, J. Wiley and Sons, New York, 1984.

[2] Anderson D.A., Aitkin M.A., Variance component models with binary response: interviewer variability, J. R. Stat. Soc. B 47 (1985) 203–210.

[3] Bertrand J.K., Benyshek L.L., Variance and covariance estimates for maternally influenced beef growth traits. J. Anim. Sci. 64 (1987) 728–734.

[4] Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. B 39 (1977) 1–38.

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

[6] Foulley J.L., Heteroskedastic threshold models with applications to the analysis of calving difficulties, Interbull Bulletins 18 (1998) 3–11.

[7] Foulley J.L., Quaas R.L., Heterogeneous variances in Gaussian linear mixed models, Genet. Sel. Evol. 27 (1995) 211–228.

[8] Foulley J.L., Quaas R.L., Than d'Arnoldi C., A link function approach to heterogeneous variance components, Genet. Sel. Evol. 30 (1998) 27–43.

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

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

[11] Hartley H.O., Rao J.N.K., Maximum likelihood estimation for the mixed analysis of variance model, Biometrika 54 (1967) 93–108.

[12] Harville D.A., Maximum likelihood approaches to variance component estimation and related problems, J. Am. Stat. Assoc. 72 (1977) 320–338.

[13] Henderson C.R., Sire evaluation and genetic trends, in: Proceedings of the animal breeding and genetics symposium in honor of Dr J Lush, American society of animal science-American dairy science association, Champaign, 1973, pp. 10-41.

[14] Henderson C.R., Applications of linear models in animal breeding, University of Guelph, Guelph, 1984.

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

[16] Laird N.M., Ware J.H., Random effects models for longitudinal data, Biometrics 38 (1982) 963–974.

[17] Lindström M.J., Bates D.M., Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data, J. Am. Stat. Assoc. 83 (1988) 1014–1022.

[18] Littell R.C., Milliken G.A., Stroup W.W., Wolfinger R.D., SAS System for mixed models, SAS Institute Inc, Cary, NC, USA, 1996.

[19] Little R.J.A., Rubin D.B., Statistical analysis with missing data, J. Wiley and Sons, New York, 1977.

[20] Liu C., Rubin D.B., The ECME algorithm: a simple extension of EM and ECM with fast monotone convergence, Biometrika 81 (1994) 633–648.

[21] Liu C., Rubin D.B.,Wu Y.N., Parameter expansion to accelerate EM: The PX- EM algorithm, Biometrika 85 (1998) 755–770.

[22] Longford N.T., Random coefficient models, Clarendon Press, Oxford, 1993.

[23] Meng X.L., van Dyk D.A., The EM algorithm - an old song sung to a fast new tune (with discussion), J. R. Stat. Soc. B 59 (1997) 511–567.

[24] Meng X.L., van Dyk D.A., Fast EM-type implementations for mixed effects models, J. R. Stat. Soc. B 60 (1998) 559–578.

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

[26] Misztal I., Comparison of computing properties of derivative and derivative-free algorithms in variance component estimation by REML, J. Anim. Breed. Genet. 111 (1994) 346–352.

[27] Patterson H.D., Thompson R., Recovery of interblock information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[28] Pothoff R.F., Roy S.N., A generalized multivariate analysis of variance model useful especially for growth curve problems, Biometrika 51 (1964) 313–326.

[29] Quaas R.L., REML Note book, Mimeo, Cornell University, Ithaca, New York, 1992.

[30] Schaeffer L.R., Dekkers J.C.M., Random regressions in animal models for test-day production in dairy cattle, in: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production 18 (1994) 443–446.

[31] Searle S.R., Matrix algebra useful for statistics, John Wiley & Sons, New York, 1982.

[32] van Dyk D.A., Fitting mixed-effects models using efficient EM-type algorithms, J. Comp. Graph. Stat. (2000) accepted.

[33] Verbeke G., Molenberghs G., Linear mixed models in practice, Springer Verlag, New York, 1997.

[34] Vonesh E.F., Carter R.L., Mixed-effects non linear regression for unbalanced repeated measures, Biometrics 48 (1992) 1–17.

[35] Wolfinger R.D., Tobias R.D., Joint estimation of location, dispersion, and random effects in robust design, Technometrics 40 (1998) 62–71.

Appendix A

Explicit expression of the system $\mathbf{F}(\text{vec}\,\boldsymbol{\alpha}') = \mathbf{h}$.

The system to be solved has the general form

$$\sum_{m=1}^{K}\sum_{n=1}^{K} f_{kl,mn}^{[t]}\alpha_{mn}^{[t+1]} = h_{kl}^{[t]} \quad \text{for} \quad k,l = 1,2,\ldots,K$$

where

$$f_{kl,mn}^{[t]} = \text{tr}\,[\mathbf{Z}_k'\mathbf{H}^{-1}\mathbf{Z}_m E(\mathbf{u}_n\mathbf{u}_l'|\mathbf{y},\boldsymbol{\Gamma}^{[t]})]$$

$$h_{kl}^{[t]} = \text{tr}\left\{\mathbf{Z}_k'\mathbf{H}^{-1}E[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\mathbf{u}_l'|\mathbf{y},\boldsymbol{\Gamma}^{[t]}]\right\}$$

Let $\mathbf{T}_{kl} = \mathbf{Z}_k'\mathbf{H}^{-1}\mathbf{Z}_l$, and $\mathbf{v}_{k_{(q\times 1)}} = \mathbf{Z}_k'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}')$ and $E_c(.)$ designate a conditional expectation given $\mathbf{y},\boldsymbol{\Gamma}^{[t]}$, the left hand side wich is symmetric can be expressed, for $K=2$, as:

	11	12	21	22
11	$E_c(\mathbf{u}_1'\mathbf{T}_{11}\mathbf{u}_1)$	$E_c(\mathbf{u}_1'\mathbf{T}_{11}\mathbf{u}_2)$	$E_c(\mathbf{u}_1'\mathbf{T}_{12}\mathbf{u}_1)$	$E_c(\mathbf{u}_1'\mathbf{T}_{12}\mathbf{u}_2)$
12		$E_c(\mathbf{u}_2'\mathbf{T}_{11}\mathbf{u}_2)$	$E_c(\mathbf{u}_2'\mathbf{T}_{12}\mathbf{u}_1)$	$E_c(\mathbf{u}_2'\mathbf{T}_{12}\mathbf{u}_2)$
21			$E_c(\mathbf{u}_1'\mathbf{T}_{22}\mathbf{u}_1)$	$E_c(\mathbf{u}_1'\mathbf{T}_{22}\mathbf{u}_2)$
22				$E_c(\mathbf{u}_2'\mathbf{T}_{22}\mathbf{u}_2)$

and the right hand side as:

11	12	21	22
$E_c(\mathbf{u}_1'\mathbf{v}_1)$	$E_c(\mathbf{u}_2'\mathbf{v}_1)$	$E_c(\mathbf{u}_1'\mathbf{v}_2)$	$E_c(\mathbf{u}_2'\mathbf{v}_2)$

Appendix B

B1. Data sets for sire and maternal grandsire models.

No	Environment		Genetic factors		Calving data I			Calving data II			
	a	b	s	t	[1]	[2]	[3]	[1]	[2]	[3]	[4]
1	1	1	1	4	23	27	13	8	6	7	0
2	1	2	1	4	14	21	22	1	13	4	1
3	1	1	1	7	20	16	6	6	6	1	1
4	1	2	1	7	13	5	3	2	5	0	0
5	1	2	1	8	9	5	4	0	4	2	0
6	1	1	2	6	6	18	12	0	4	0	8
7	1	1	2	7	13	5	3	4	2	1	0
8	1	2	2	8	39	10	5	4	13	1	10
9	2	1	2	8	51	26	4	7	6	5	9
10	2	2	3	5	8	8	5	0	0	3	4
11	2	1	3	5	16	24	17	0	5	2	12
12	2	2	3	5	14	13	3	0	2	2	6
13	2	3	3	2	60	37	14	10	1	1	25
14	2	2	3	2	22	13	4	1	1	4	7
15	2	3	4	7	27	9	3	10	3	0	0
16	2	3	4	8	9	5	4	3	1	0	2
17	2	2	4	9	30	19	8	0	0	6	13
18	2	2	4	5	23	11	2	6	5	1	0

Calving data I: Foulley [6]. Calving performance scored according to an increasing level of dystocia as factors of a = sex, b = age of dam, s = sire of calf and t = maternal grandsire of calf.

Calving data II: Same design structure with 4 score values, smaller subclass numbers and simulated data.

a,b,s,t: stand for factors a, b (fixed) and sire and maternal grandsire (random) respectively. Non-zero elements $(i, j) = (j, i)$ of the numerator relationship matrix are: $(1, 2) = (8, 9) = 1/4$; $(1, 5) = (2, 5) = (3, 7) = (4, 6) = (8, 10) = (9, 10) = 1/2$; $(i, i) = 1$, for any $i = 1, 2, \ldots, 10$.

B2. Growth measurements in 11 girls and 16 boys (from Pothoff and Roy [28]; Little and Rubin [19]).

	Age (years)					Age (years)			
Girl	8	10	12	14	Boy	8	10	12	14
1	210	200	215	230	1	260	250	290	310
2	210	215	240	255	2	215		230	265
3	205		245	260	3	230	225	240	275
4	235	245	250	265	4	255	275	265	270
5	215	230	225	235	5	200		225	260
6	200		210	225	6	245	255	270	285
7	215	225	230	250	7	220	220	245	265
8	230	230	235	240	8	240	215	245	255
9	200		220	215	9	230	205	310	260
10	165		190	195	10	275	280	310	315
11	245	250	280	280	11	230	230	235	250
					12	215		240	280
					13	170		260	295
					14	225	255	255	260
					15	230	245	260	300
					16	220		235	250

Distance from the centre of the pituary to the pteryomaxillary fissure (unit 10^{-4}m).

B3: Ultrafiltration data set (from [18, 33]).

#	QB	TMP	UFR	#	QB	TMP	UFR	#	QB	TMP	UFR	#	QB	TMP	UFR
	2	240	645		2	260	3660		3	255	3885		3	235	1170
	2	505	20115		2	500	16950		3	500	19155		3	485	17685
	2	995	38460		2	1020	36090		3	980	37650		3	1025	39705
1	2	1485	44985	6	2	1490	42630	11	3	1490	47895	16	3	1515	52680
	2	2020	51765		2	1990	46470		3	2015	54495		3	1990	61800
	2	2495	46575		2	2480	46275		3	2510	53175		3	2510	61485
	2	2970	40815		2	2995	43980		3	2980	59355		3	3020	61425
	2	240	3720		2	305	9825		3	280	5715		3	285	1500
	2	540	18885		2	505	21630		3	505	20505		3	520	15405
	2	995	34695		2	980	42270		3	1000	39405		3	1005	32520
2	2	1475	40305	7	2	1505	50280	12	3	1490	50100	17	3	1500	42435
	2	2000	44475		2	2005	45510		3	2000	55155		3	1985	48570
	2	2500	42435		2	2505	44250		3	2505	61185		3	2490	53685
	2	3010	44655		2	2990	42300		3	3020	50715		3	2995	53655
	2	245	2985		2	305	9480		3	355	10410		3	295	6420
	2	480	17700		2	505	21750		3	480	19320		3	515	20250
	2	1010	35295		2	995	37230		3	1025	43770		3	1010	43050
3	2	1505	41955	8	2	1500	44430	13	3	1500	51225	18	3	1480	58110
	2	2000	47610		2	1990	42165		3	1990	58095		3	2000	61995
	2	2515	44730		2	2480	43065		3	2500	54090		3	2480	60915
	2	2970	46035		2	3000	36615		3	3005	62010		3	3005	63600
	2	255	3930		2	250	1560		3	235	3600		3	290	4050
	2	495	19830		2	495	16650		3	480	20490		3	495	16590
	2	995	40425		2	1000	34530		3	1010	41880		3	1015	40515
4	2	1480	52260	9	2	1500	43815	14	3	1490	49995	19	3	1520	52845
	2	1995	49395		2	1965	48495		3	1990	57675		3	2020	60435
	2	2490	45975		2	2485	47520		3	2480	62475		3	2500	64830
	2	3030	41910		2	2980	41640		3	3005	62145		3	2975	63825
	2	255	3210		2	235	1230		3	260	1890		3	400	10935
	2	515	17700		2	505	15375		3	515	18510		3	470	13470
	2	1000	32490		2	1020	32835		3	970	37215		3	1010	35355
5	2	1505	42330	10	2	1475	37830	15	3	1505	52350	20	3	1515	45345
	2	2020	45735		2	1970	40590		3	1990	60915		3	1980	49440
	2	2490	47850		2	2480	32550		3	2500	62985		3	2510	53625
	2	3010	48045		2	3000	34305		3	2995	64770		3	3000	56430

#: = Dializer identification; QB$\times 10^{-2}$ = Blood flow rate; TMP$\times 10^{3}$ = Transmembrane pressure; UFR$\times 10^{3}$ = Ultrafiltration rate.