

# Estimation in mixed models through Three Step Minimization

DÁRIO FERREIRA<sup>1</sup>, SANDRA S. FERREIRA<sup>1</sup>, CÉLIA NUNES<sup>1</sup>  
AND JOÃO T. MEXIA<sup>2</sup>

<sup>1</sup> Department of Mathematics, University of Beira Interior, Covilhã, Portugal;

<sup>2</sup> Department of Mathematics, Faculty of Science and Technology, New University of Lisbon, Monte da Caparica, Portugal

*The aim of this article is to present an estimation procedure for both fixed effects and variance components in linear mixed models. This procedure consists of a maximum likelihood method which we call Three Step Minimization, TSM. The major contribution of this method is that when variances tend to be null standard algorithms behave badly, unlike the TSM method, which uses a grid search algorithm in a compact set. A numerical application with real and simulated data is provided.*

**Keywords** Mixed models; Inference; Maximum likelihood; Variance components; Newton-Raphson.

**Mathematics Subject Classification** 60E05; 62F25.

## 1 Introduction

The estimation problem of variance components in the case of balanced designs in mixed models has been extensively treated, see for example Khuri and Sahai (1985), Searle (1971) and Searle (1995). Inference becomes less elegant when the orthogonality conditions fail to hold.

Prior to the early 1990s, most applications used some version of analysis of variance (ANOVA). The sums of squares for an ANOVA table are easily calculated and are unambiguous, see for example LaMotte (1973) and Vanleeuwen et al (1999). The early development of ANOVA estimation of variance components for unbalanced data are attributed to Henderson (1953). Nowadays, with the development of computing facilities, the use of maximum likelihood (ML) methods has sparked considerable interest since, despite the methodology for ML estimation is simple, its implementation is mathematically intense. The likelihood function is maximized

---

Received September 2, 2013

Address correspondence to Dário Ferreira, Department of Mathematics, University of Beira Interior, Avenida Marquês D' Ávila e Bolama, 6200 Covilhã, Portugal E-mail: dario@ubi.pt

over the parameter space under non-negative definite estimates of the variance matrices where data is assumed to be normally distributed, see Khuri (1985). Under certain regularity conditions, ML estimators of variance components are consistent, asymptotically efficient and asymptotically normal, see Miller (1977).

Jennrich and Schluchter (1985) and Fairclough and Helms (1986) compared performances of the Newton-Raphson, NR, the Fisher of Scoring method, FS, and the EM algorithm. Lindstrom and Bates (1988) provided arguments favoring the use of the NR method. Based on some starting values for the parameters, procedures based on NR method iteratively update the estimates until convergence is achieved, see Ypma (1995). The difference for the FS is that the FS replaces the observed second derivatives with its expectation. This method was introduced by Fisher (1925). The NR approach is usually very good if the initial estimate is close to the value of the parameter. But if it is not close enough the iterative process may converge too slowly, may not converge at all or may converge to parameter values on or outside the boundary of the parameter space. This problem might be circumvented by specifying better starting values or by using other numerical procedures, see Hartley and Rao (1967) or Stern and Welsh (2000). Problems concerning the estimation of variance components in mixed effects models were extensively discussed in the monograph by Demidenko (2013).

In this paper we present a maximum likelihood method for the estimation of both fixed effects and variance components in balanced and unbalanced linear mixed models, which we call Three Step Minimization, TSM. With this approach we first profile out the fixed parameters. Then, after a re-parametrization we switch to using ratio parameters profiling out the variance components. Finally we use a grid search algorithm to carry out a minimization, choosing those ratio parameters that minimize the transformed log likelihood function.

The main advantage of our method is that the grid search is robust and can easily identify variances that are zero or tend to zero. To the contrary, standard algorithms, such as NR or FS, behave badly in this situation. Besides this, TSM is very easy to implement in R software, which is free and available on the internet, and it does not need any starting value, since the grid search will be restricted to the interval  $[0, 1]$ . Moreover, our estimates may be used as excellent starting values for the NR method.

This article is structured as follows. In the next section we make a brief presentation of linear mixed models. Then we explain how to estimate the parameters through the TSM method. The fourth section provides a numerical comparison of our approach with the NR and FS. Namely we present a numerical example applied to simulated and real data. We used designs with small number of replicates and with one variance component very close to zero, since these are the type of designs we are interested in. Apart from the example in the application with real data this type of design may be used, for example in medicine, in which patients suffer from infrequent pathologies. Finally in the last section we present some comments.

## 2 Model

Let  $Y = (Y_1, \dots, Y_n)^\top$  describe the observation vector in the linear mixed model

$$Y = X\beta + \sum_{l=1}^{k-1} Z_l b_l + e, \quad (1)$$

where  $X$  and  $Z_l$  are known  $n \times m$  and  $n \times c_l$  matrices respectively,  $X$  has full rank,  $\beta \in \mathbb{R}^m$  is fixed and the random effects  $b_l$  and  $e$  are assumed to be normally distributed and independent. We put

$$b_l \sim N(0, \sigma_l^2 I_{c_l}), \quad (2)$$

$l = 1, \dots, k-1$  and

$$e \sim N(0, \sigma_k^2 I_n). \quad (3)$$

Therefore

$$E(Y) = X\beta \quad (4)$$

and

$$V(\sigma^2) = \text{Var}(Y) = \sum_{l=1}^k \sigma_l^2 M_l, \quad (5)$$

where  $M_l = Z_l Z_l^\top$ ,  $l = 1, \dots, k-1$ ,  $M_k = I_n$  and  $\sigma^2 = (\sigma_1^2, \dots, \sigma_k^2)^\top$ . Thus we have

$$Y \sim N(X\beta, V(\sigma^2)) \quad (6)$$

which has density

$$n(Y|X\beta, V(\sigma^2)) = \frac{e^{-\frac{1}{2}(Y-X\beta)^\top V(\sigma^2)^{-1}(Y-X\beta)}}{(2\pi)^{n/2} \sqrt{|V(\sigma^2)|}}. \quad (7)$$

Having thus the log-likelihood function

$$l(\beta, \sigma^2 | Y) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |V(\sigma^2)| - \frac{1}{2} (Y - X\beta)^\top V(\sigma^2)^{-1} (Y - X\beta), \quad (8)$$

with  $|A|$  describing the determinant of matrix  $A$ .

In the next section we will show how to estimate the parameter vector  $\theta = (\beta, \sigma^2)^\top$  using our approach. The description of the NR method may be seen in Demidenko (2013).

### 3 Three Step Minimization

Maximizing (8) is equivalent to minimizing

$$l^*(\beta, \sigma^2 | Y) = \log |V(\sigma^2)| + (Y - X\beta)^\top V(\sigma^2)^{-1} (Y - X\beta). \quad (9)$$

We will minimize (9) in three steps. In the first step we will replace the goal function by another which depends only on the variance components. In the second step we replace the goal function depending on the variance components by another which depends on ratio parameters. The third step consists of a numerical minimization.

- First step:

The minimum of  $l^*(\beta, \sigma^2 | Y)$  for a fixed  $\sigma^2$  is obtained for the corresponding least squares estimator

$$\tilde{\beta} = \left( X^\top V(\sigma^2)^{-1} X \right)^{-1} X^\top V(\sigma^2)^{-1} Y. \quad (10)$$

Thus

$$Y - X\tilde{\beta} = (I_n - X(X^\top V(\sigma^2)^{-1} X)^{-1} X^\top V(\sigma^2)^{-1}) Y, \quad (11)$$

and moreover

$$(Y - X\tilde{\beta})^\top V(\sigma^2)^{-1} (Y - X\tilde{\beta}) = Y^\top P Y, \quad (12)$$

where  $P = V(\sigma^2)^{-1} - V(\sigma^2)^{-1} X (X^\top V(\sigma^2)^{-1} X)^{-1} X^\top V(\sigma^2)^{-1}$ .

So the minimum of  $l^*(\beta, \sigma^2 | Y)$  for a fixed  $\sigma^2$  will be given by the profile log likelihood

$$l^{**}(\sigma^2 | Y) = \log |V(\sigma^2)| + Y^\top P Y, \quad (13)$$

see Pinheiro and Bates (2000). We should point out that now, after profiling out the fixed parameters, the number of variables to be estimated has been reduced substantially since we only have to consider  $\sigma^2$ .

- Second step:

In the second step we will consider ratio parameters  $\gamma_l$ , with

$$\gamma_l = \frac{\sigma_l^2}{\rho}, \quad l = 1, \dots, k, \quad (14)$$

where  $\rho = \sum_{l=1}^k \sigma_l^2$ , profiling out the parameter corresponding to the sum of the variance parameters.

Let us assume that  $\gamma = (\gamma_1, \dots, \gamma_k)$ . Now

$$V(\sigma^2) = \sum_{l=1}^k \sigma_l^2 M_l = \rho \sum_{l=1}^k \gamma_l M_l = \rho V(\gamma), \quad (15)$$

with

$$V(\gamma) = \sum_{l=1}^k \gamma_l M_l, \quad l = 1, \dots, k. \quad (16)$$

So  $l^{**}(\sigma^2 | Y)$  may be rewritten as

$$\begin{aligned} l^{**}(\rho; \gamma | Y) &= \log(\rho^n | V(\gamma)) + \\ &+ Y^\top \left[ \frac{1}{\rho} V(\gamma)^{-1} - \frac{1}{\rho} V(\gamma)^{-1} X (X^\top \frac{1}{\rho} V(\gamma)^{-1} X)^{-1} X^\top \frac{1}{\rho} V(\gamma)^{-1} \right] Y \\ &= n \log \rho + \log(|V(\gamma)|) \\ &+ \frac{1}{\rho} Y^\top [V(\gamma)^{-1} - V(\gamma)^{-1} X (X^\top V(\gamma)^{-1} X)^{-1} X^\top V(\gamma)^{-1}] Y. \end{aligned} \quad (17)$$

Thus

$$\begin{aligned} \frac{\partial l^{**}(\rho; \gamma | Y)}{\partial \rho} &= \frac{n}{\rho} - \frac{1}{\rho^2} Y^\top [V(\gamma)^{-1} \\ &- V(\gamma)^{-1} X (X^\top V(\gamma)^{-1} X)^{-1} X^\top V(\gamma)^{-1}] Y. \end{aligned} \quad (18)$$

So, considering

$$Q = Y^\top [V(\gamma)^{-1} - V(\gamma)^{-1} X (X^\top V(\gamma)^{-1} X)^{-1} X^\top V(\gamma)^{-1}] Y, \quad (19)$$

the minimum for the given  $\gamma$  is attained with

$$\rho = \frac{Q}{n}. \quad (20)$$

This minimum will be

$$l(\gamma | Y) = n \log\left(\frac{Q}{n}\right) + \log(|V(\gamma)|) + n. \quad (21)$$

• Third step:

In this step we will minimize the variance-profile log-likelihood function in (21). To do that we take the ratio parameters defined in (14) and use a stochastic or a grid search algorithm to carry out a numerical minimization, choosing those  $\gamma_l, l = 1, \dots, k$ , that minimize  $l(\gamma | Y)$ , and complete the adjustment. Note that the  $\gamma_l, l = 1, \dots, k$  will vary in the compact set

$$\mathcal{D} = \{\gamma_1, \dots, \gamma_k\}, \quad (22)$$

where  $0 \leq \gamma_l \leq 1, l = 1, \dots, k$  and  $\sum_{l=1}^k \gamma_l = 1$ . So this minimization is under constraint of these conditions. In the numerical comparisons we used a grid search algorithm. A brief description of it may be found in the next section. Moreover, the scheme of the grid search algorithm may be seen at

<http://dar364.wix.com/dario#algorithms/c1oi>.

Finally there will be

$$\tilde{\sigma}_l^2 = \rho \tilde{\gamma}_l, \quad l = 1, \dots, k. \quad (23)$$

To estimate the remaining parameters we may now use  $\tilde{\sigma}^2$  to obtain  $V(\tilde{\sigma}^2)$ , the estimator of  $V(\sigma^2)$ , in (5), and then use it in (10) providing

$$\tilde{\beta} = \left( X^\top V(\tilde{\sigma}^2)^{-1} X \right)^{-1} X^\top V(\tilde{\sigma}^2)^{-1} Y. \quad (24)$$

Thus, using the TSM method we are able to estimate the variance components and the fixed effects for an arbitrary model.

## 4 Numerical application

### 4.1 Numerical comparison using simulated data

In order to compare the estimates of the variance components obtained through the TSM, and the ones obtained through the NR and FS, using simulated data and with one variance component close to zero, we considered 7 unbalanced models,  $s = 1, \dots, 7$ . Each model has a three level fixed effects factor,  $A$ , that crosses with a two level fixed effects factor,  $B$ , which nests an  $r$  level random effects factor,  $C$ , where  $r = s + 1$ . So, the 1-st model is  $A^3 \times (B^2 \supset C^2)$ , while the 7-th model is  $A^3 \times (B^2 \supset C^8)$ . We used these models because they have a combination of nested and crossed effects and are widely used, see for example Khuri et al (1998). Table 1 shows how the data are distributed in the 7-th model, that is, in the case where  $C$  factor has 8 levels. The remaining models are obtained by removing levels of the  $C$  factor. The table contains dots instead of values because the data were simulated 10000 times.

In Section 4.2 we will use a very similar design to the first one ( $s = 1$ ) presented in this section. We thought of using this design because so we are replicating the design described in the application to real data. When  $s=1$ , the design is almost the same. Thus, this design could be used for the investigation described in the application to real data but varying the number of clones between 2 ( $s=1$ ) and 8 ( $s=7$ ).

The models for the observation vectors  $Y_s^\top = (Y_1, \dots, Y_{n_s})$ ,  $s = 1, \dots, 7$ , are given by

$$Y_s = X_s \beta + Z_{1,s} b_{1,s} + Z_{2,s} b_{2,s} + e_s \quad (25)$$

where  $\beta$  is a  $12 \times 1$  fixed vector,

$$\begin{cases} b_{1,s} \sim N(0, \sigma_{1,s}^2 I_{6(s+1)}) \\ b_{2,s} \sim N(0, \sigma_{2,s}^2 I_{2(s+1)}) \\ e_s \sim N(0, \sigma_{3,s}^2 I_{n_s}) \end{cases}, \quad (26)$$

$s = 1, \dots, 7$ , are independent and  $X_s$ ,  $Z_{1,s}$  and  $Z_{2,s}$  are known design matrices.

Let

$$(\sigma_s^2)^\top = (0.05, 0.5, 1) \quad (27)$$

Table 1: Distribution of observations by factors for the 7-th model.

|       | $B_1$  | $B_2$  |
|-------|--|--|
|       | $C_1 C_2 C_3 C_4 C_5 C_6 C_7 C_8$  | $C_1 C_2 C_3 C_4 C_5 C_6 C_7 C_8$  |
| $A_1$ | ● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ● |
| $A_2$ | ● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ● |
| $A_3$ | ● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ● | ● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ●<br>● ● ● ● ● ● ● ● |

be the vector with the assumed numbers for the variance components and

$$\beta_{\cdot}^{\top} = (\beta_1, \dots, \beta_{12}), \quad (28)$$

where  $\beta_j = 1$ ,  $j = 1, \dots, 12$ , be the vector with the assumed numbers for the fixed effects. Note that one of the assumed numbers for the variance components is 0.05.

Hereinafter we used the R software to proceed as follows  $N = 10000$  times:

We simulated  $Y_s$ ,  $s = 1, \dots, 7$  with the above assumed numbers for the parameters. Then despite the observation vectors and all parameters are known, we assumed that we only knew  $Y_s$ ,  $X_s$ ,  $Z_{1,s}$  and  $Z_{2,s}$ ,  $s = 1, \dots, 7$  and used both methods to estimate the parameters and compare the obtained estimates with the assumed numbers in (27) and (28).

To estimate the parameters using the TSM method we minimized the function in (21) with respect to  $\gamma_s$ ,  $s = 1, \dots, 7$ . To do that we considered  $\gamma_s = (\gamma_{1,s}, \gamma_{2,s}, \gamma_{3,s})$  and used a grid search algorithm, whose scheme is presented at [http : //dar364.wix.com/dario#algorithms/c1oi](http://dar364.wix.com/dario#algorithms/c1oi). The programmed algorithm for  $s = 1$  is also available. The algorithm makes vary  $\gamma_{l,s}$ ,  $l = 1, 2, 3$ ,  $s = 1, \dots, 7$ , such that

$$\begin{cases} \gamma_{1,s} = 0.01 \times k \\ \gamma_{2,s} = 0.01 \times i \times (1 - \gamma_{1,s}) \\ \gamma_{3,s} = 1 - (\gamma_{1,s} + \gamma_{2,s}) \end{cases}, \quad (29)$$

with  $k = 0, \dots, 100$  and  $i = 0, \dots, 100$ , thus having

$$\begin{cases} 0 \leq \gamma_{l,s} \leq 1, \quad l = 1, 2, 3 \\ \gamma_{1,s} + \gamma_{2,s} + \gamma_{3,s} = 1 \end{cases}, \quad (30)$$

$s = 1, \dots, 7$ . Proceeding that way it calculated the value of  $l(\gamma|Y)$ , in (21), for each  $\gamma_s$  and gave us those that minimize  $l(\gamma|Y)$ . Next we obtained the  $\tilde{\sigma}_s^2$ ,  $s = 1, \dots, 7$ ,

from (23) and the  $\tilde{\beta}_s$ ,  $s = 1, \dots, 7$ , from (24). Note that there is no need to search for a starting value, since our search is restricted to the interval  $[0, 1]$ .

The idea behind the NR method is to approximate a given function in each iteration by a quadratic function and then move to the minimum of this quadratic, see Ypma (1995). The quadratic approximation in a suitable neighborhood of a given point,  $\theta$ , with components  $\theta_1, \dots, \theta_w$ , is given by a second order Taylor expansion

$$l(\theta) \approx l(\theta_0) + (\theta - \theta_0)^\top \frac{\partial l(\theta)}{\partial \theta} \Big|_{\theta_0} + \frac{1}{2} (\theta - \theta_0)^\top \left( \frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta_0} \right) (\theta - \theta_0). \quad (31)$$

So, to estimate the parameters using the NR method we minimized the function in (31). To do that we used an algorithm whose scheme is presented at <http://dar364.wix.com/darioalgorithms/c1oi>. For the maximum difference between two successive iterations we considered  $\epsilon = 0.01$ . We used this  $\epsilon$  to ensure that the conditions were similar for both methods.

To apply the NR method we needed to consider some starting values. We could use an algorithm to select them but as we wanted to study the behavior of the NR with smaller and larger starting values we used the values presented in Table 2. If the starting value is smaller than the assumed number,  $AN$ , in (27), we call the NR method NR-L. Otherwise we call it NR-R. For example, if  $AN = 1$ , the starting values for NR-L and NR-R methods will be 0.75 and 1.25 respectively.

Table 2: Starting values for NR-L and NR-R.

| $AN$ : | 0.05  | 0.5  | 1    |
|--------|-------|------|------|
| NR-L   | 0.001 | 0.25 | 0.75 |
| NR-R   | 0.25  | 0.75 | 1.25 |

To compare the TSM with the NR let

$$\bar{\sigma}_s^2 = \frac{1}{N} \sum_{j=1}^N \tilde{\sigma}_{s,j}^2, \quad s = 1, \dots, 7, \quad (32)$$

and

$$\bar{\beta}_s = \frac{1}{N} \sum_{j=1}^N \tilde{\beta}_{s,j}, \quad s = 1, \dots, 7. \quad (33)$$

Moreover let

$$\tilde{V}_{s,\cdot} = \frac{1}{N-1} \left( \sum_{j=1}^N \tilde{\sigma}_{s,j}^2 (\tilde{\sigma}_{s,j}^2)^\top - N \bar{\sigma}_s^2 (\bar{\sigma}_s^2)^\top \right), \quad s = 1, \dots, 7, \quad (34)$$

and

$$\tilde{V}_{s,\bullet} = \frac{1}{N-1} \left( \sum_{j=1}^N \tilde{\beta}_{s,j} \tilde{\beta}_{s,j}^\top - N \bar{\beta}_s \bar{\beta}_s^\top \right), \quad s = 1, \dots, 7. \quad (35)$$



Table 3 shows the obtained values for

$$d_{1,s} = \|\bar{\sigma}_s^2 - \sigma_s^2\|, \quad s = 1, \dots, 7, \quad (36)$$

and Table 4 shows the obtained values for the trace of matrix  $\tilde{V}_{s,\cdot}$ ,  $\text{tr}(\tilde{V}_{s,\cdot})$ ,  $s = 1, \dots, 7$ , respectively.

Table 3: Obtained values for  $d_{1,s}$ ,  $s = 1, \dots, 7$ .

| method | 1      | 2      | 3      | 4      | 5      | 6      | 7      |
|--------|--------|--------|--------|--------|--------|--------|--------|
| TM     | 0.1534 | 0.3854 | 0.1026 | 0.2728 | 0.2003 | 0.1085 | 0.2348 |
| NR-L   | 2.7087 | 0.4634 | a)     | 0.2822 | 0.2624 | 0.3561 | 0.3557 |
| NR-R   | a)     | 1.7649 | 0.3760 | 0.3282 | 0.1538 | 4.9909 | 2.2241 |

a) Function in (31) did not converge.

Table 4: Values of  $\text{tr}(\tilde{V}_{s,\cdot})$ ,  $s = 1, \dots, 7$ .

|      | 1      | 2       | 3       | 4       | 5      | 6      | 7      |
|------|--------|---------|---------|---------|--------|--------|--------|
| TM   | 0.0341 | 0.0536  | 0.0210  | 0.0421  | 0.0209 | 0.0981 | 0.0172 |
| NR-L | 0.9760 | 0.0104  | a)      | 0.0708  | 0.0002 | 0.1029 | 0.0772 |
| NR-R | a)     | 18.2697 | 14.4334 | 17.0105 | 0.0221 | 1.4981 | 2.0143 |

a) Function in (31) did not converge.

It may be seen that in general the TSM method gave more accurate values with smaller variation. Moreover, NR-L did not converge when  $s = 3$  and NR-R did not converge when  $s = 1$ . Note that both algorithms may be improved. Perhaps using typical algorithms to select starting values would provide a more realistic representation of NR implementation and performance, while for the TSM we could reduce the net in the compact set. Or alternatively we could have used any other grid search algorithm or even a stochastic algorithm. However our intention was to show that TSM method finds good estimates when the ML solution is on the boundary of the minimization domain and does not require starting values. Moreover, once we have the estimates obtained by the TSM these could also be improved by using them as starting values for the NR. We will see an application of this in Section 4.2.

Table 5 shows the obtained values for

$$d_{2,s} = \|\bar{\beta}_s - \beta_s\|, \quad s = 1, \dots, 7. \quad (37)$$

and Table 4 shows the obtained values for the trace of matrix  $\tilde{V}_{h,\cdot}$ ,  $\text{tr}(\tilde{V}_{h,\cdot})$ ,  $s = 1, \dots, 7$ .

It may be seen that the values of  $d_{2,s}$  and  $\text{tr}(\tilde{V}_{h,\cdot})$ ,  $s = 1, \dots, 7$  are very similar. Thus, we may conclude that in this case the method used to estimate the variance components does not influence the estimates obtained for  $\beta$ .

Table 7 shows the sum of the time spent (10.000 cases), in seconds, in obtaining the estimates of the variance components and the fixed effects. It may be seen that

Table 5: Obtained values for  $d_{2,s}$ ,  $s = 1, \dots, 7$ .

| method | 1      | 2      | 3      | 4      | 5      | 6      | 7      |
|--------|--------|--------|--------|--------|--------|--------|--------|
| TM     | 2.1284 | 2.0858 | 2.1344 | 2.0424 | 2.0850 | 2.1666 | 2.0855 |
| NR-L   | 2.1355 | 2.0807 | a)     | 2.0428 | 2.0844 | 2.1666 | 2.0824 |
| NR-R   | a)     | 2.0843 | 2.1407 | 2.0449 | 2.0965 | 2.1676 | 2.0962 |

a) Function in (31) did not converge.

Table 6: Values of  $\text{tr}(\tilde{V}_{s,\cdot\cdot})$ ,  $s = 1, \dots, 7$ .

|      | 1      | 2      | 3      | 4      | 5      | 6      | 7      |
|------|--------|--------|--------|--------|--------|--------|--------|
| TM   | 0.4209 | 0.7895 | 0.7967 | 0.6791 | 0.1291 | 0.6820 | 0.3265 |
| NR-L | 0.4275 | 0.7870 | a)     | 0.6871 | 0.1133 | 0.6807 | 0.3213 |
| NR-R | a)     | 0.7870 | 0.7869 | 0.6786 | 0.1284 | 0.6808 | 0.3571 |

a) Function in (31) did not converge.

Table 7: Sum of the time spent, in seconds, in obtaining the values of  $\tilde{\sigma}^2$  and  $\tilde{\beta}$ .

| method | 1      | 2      | 3      | 4      | 5      | 6      | 7      |
|--------|--------|--------|--------|--------|--------|--------|--------|
| TM     | 20.05  | 25.43  | 29.72  | 34.97  | 45.20  | 61.79  | 72.07  |
| NR-L   | 223.05 | 183.52 | a)     | 178.35 | 358.72 | 365.75 | 360.63 |
| NR-R   | a)     | 332.93 | 346.71 | 324.38 | 356.50 | 355.08 | 353.01 |

a) Function in (31) did not converge.

TSM method took less time than the NR-L and NR-R. However, as the values are all in fractions of seconds we conclude that time is trivial.

Besides this comparison we also estimated the variance components using SPSS software through the FS. To do that we used one of the simulated observation vectors per model. The used observation vectors are available at <http://dar364.wix.com/dario#algorithms/c1oi>. We used SPSS instead of R software to compare the estimates obtained by our algorithms with estimates obtained by another software. The obtained estimates, as well as  $d_{1,s}$ ,  $s = 1, \dots, 7$ , are presented in Table 8.

Table 8: Obtained values for  $d_{1,s}$ ,  $s = 1, \dots, 7$ , using SPSS software through the FS.

|           | 1      | 2      | 3      | 4      | 5      | 6      | 7      |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| $d_{1,s}$ | 0.3693 | 0.4664 | 0.2522 | 0.5178 | 0.3692 | 0.2630 | 0.2052 |

Comparing values in Table 3 and Table 8 we see that, excluding the case where  $s = 2$ ,  $s = 4$  and  $s = 5$ , SPSS through the FS gave more accurate estimates than the R software through our algorithms for NR-L and NR-R. But excluding the case where  $s = 7$ , R software through the TSM gave more accurate estimates than the SPSS through the FS.

## 4.2 Application to real data

We considered an experiment in which two groups of two clones, grown side by side, of the caste Touriga Nacional. The origins of the two groups were distinct. As before there is a fixed effects factor (location) that crosses with another fixed effects factor (origin) which nests a random effects factor (clone). The yields, in Kg, are presented in Table 9. These data were presented in Fonseca et al. (2003) in an application to a grapevine experiment. Here we introduced imbalance by deleting some observations from the original table. The yields are presented in Table 9.

Table 9: Yields in Kg.

| Location | Origin 1 |         | Origin 2 |         |
|----------|----------|---------|----------|---------|
|          | Clone 1  | Clone 2 | Clone 1  | Clone 2 |
| 1        | 3.00     | 1.00    | 1.75     | 1.05    |
|          | 1.85     | 1.10    | 2.00     | 1.25    |
| 2        | 1.80     | 1.60    | 2.00     | 2.70    |
|          | 0.70     | 1.75    | 2.55     | 2.15    |
|          | 2.50     |         | 2.65     | 2.10    |
| 3        | 1.05     | 0.75    | 1.60     | 1.60    |
|          | 1.50     | 0.65    | 1.66     | 1.10    |
|          |          | 0.90    |          |         |

The model for the observation vector  $Y$  is

$$Y = X\beta + Z_1b_1 + Z_2b_2 + e$$

where  $\beta$  is a  $12 \times 1$  fixed vector,

$$\begin{cases} b_1 \sim N(0, \sigma_1^2 I_{12}) \\ b_2 \sim N(0, \sigma_2^2 I_4) \\ e \sim N(0, \sigma_3^2 I_{28}) \end{cases},$$

are independent and  $X$ ,  $Z_1$  and  $Z_2$  are known design matrices.

Using again the R software we obtained the estimates for the variance components applying the TSM method which are

$$\begin{cases} \tilde{\sigma}_1^2 = 0.025120 \\ \tilde{\sigma}_2^2 = 0.000129 \\ \tilde{\sigma}_3^2 = 0.185249 \end{cases}.$$

Then we used these values as starting values for The NR and obtained the estimates presented in Table 10. We used  $\epsilon = 10^{-3}$ ,  $\epsilon = 10^{-4}$ ,  $\epsilon = 10^{-5}$  and  $\epsilon = 10^{-6}$ . The number of iterations is represented by  $n$ .

Table 10: Obtained estimates for the variance components using the TSM estimates as starting values for the NR.

|                      | $\epsilon = 10^{-3}$ | $\epsilon = 10^{-4}$ | $\epsilon = 10^{-5}$ | $\epsilon = 10^{-6}$ |
|----------------------|----------------------|----------------------|----------------------|----------------------|
| $\tilde{\sigma}_1^2$ | 0.025120             | 0.025120             | 0.025120             | 0.025120             |
| $\tilde{\sigma}_2^2$ | 0.000129             | 0.000129             | 0.000129             | 0.000129             |
| $\tilde{\sigma}_3^2$ | 0.184498             | 0.185326             | 0.185248             | 0.185249             |
| $n$                  | 112                  | 964                  | 27045                | 105971               |

Comparing these results from those obtained through the TSM method we see that the estimates for  $\sigma_1^2$  and  $\sigma_2^2$  obtained using the NR were identical to those obtained through the TSM, irrespective of the number of iterations. Besides this the estimate for  $\sigma_3^2$  obtained using the NR started by diverging from that obtained through the TSM, but with the increase of iterations these approached. Thus we may conclude that the estimates obtained through the TSM were validated by the NR.

We also obtained the variance components estimates through the NR and FS, using SPSS software, which are presented in Table 11. We used  $\epsilon = 10^{-8}$ .

Table 11: Obtained estimates for the variance components through the NR and FS using SPSS software.

|                      | NR       | FS       |
|----------------------|----------|----------|
| $\tilde{\sigma}_1^2$ | 0.027759 | 0.027759 |
| $\tilde{\sigma}_2^2$ | 0        | 0        |
| $\tilde{\sigma}_3^2$ | 0.183873 | 0.183873 |

It may be seen that the estimates obtained by SPSS through the NR and FS are very close to the ones obtained by the TSM and the NR with the TSM estimates as starting values. So, once again, we may conclude that the estimates obtained by our algorithms were validated by SPSS through the NR and FS.

## 5 Final Comments

We proposed a maximum likelihood method, the TSM method, for the estimation of both fixed effects and variance components in balanced and unbalanced mixed models.

A numerical comparison with the NR and FS was provided. In general the TSM method gave more accurate values than the NR. As stated before, we could use an algorithm to select the starting values for the NR but as we wanted to study the behavior of the NR with smaller and larger starting values we used the values presented in Table 2. These are for sure not the best starting values for the NR but note that the TSM method can also be improved by making the net narrower in step 3. Comparing with the FS method, the TSM method obtained more accurate results than the FS

in six of seven cases. The FS method was more accurate than the TSM for  $s = 7$ , obtaining  $d_{1,2} = 0.2052$ , while with the TSM method the result was  $d_{1,2} = 0.2348$ . Despite the added value of this comparison, we think that the major contribution of the TSM method is that when variances tend to be null, i.e. the ML solution is on the boundary of the minimization domain, standard algorithms, such as NR or FS, behave badly, unlike the TSM method, which uses a grid search in the compact set  $[0, 1]$ . In addition, TSM does not need a starting value. So, in this sense, this method proves easier and faster to apply than the Newton Raphson. Besides this, TSM is very easy to implement in R software, which is free and available on the internet.

### Acknowledgements

The authors gratefully acknowledge the reviewers for their very valuable comments.

This work was partially supported by the Center of Mathematics, University of Beira Interior through the project PEst-OE/MAT/UI0212/2014 and by CMA, Faculty of Science and Technology, New University of Lisbon, through the project PEst-OE/MAT/UI0297/2011.

### References

- Demidenko, E. (2013). *Mixed Models: Theory and Applications with R, 2nd Edition*. Wiley, New York.
- Fairclough, D. L., Helms, R. H. (1986). A mixed linear model with linear covariance structure: a sensitivity analysis of maximum likelihood estimators. *J. Statist. Comput. Simulation*. 25: 205–236.
- Fisher, R. A. (1925). *Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- Fonseca, M., Mexia, J. T., Zmyslony, R. (2003). Estimating and Testing of Variance Components: An Application to a Grapevine Experiment. *Biometrical Letters* 40(1):1–7.
- Hartley, H., Rao, J. (1967). Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model. *Biometrika*. 54(1/2): 93–108.
- Henderson, C. (1953). Estimation of variance and covariance components. *Biometrics*. 9: 226–252.
- Jennrich, R. I., Schluchter, M. D. (1985). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*. 42: 805–820.
- Khuri, A., Sahai, H. (1985). Variance components analysis: a selective literature survey. *International Statistical Review*. 53: 279–300.
- Khuri, A. (1985) *Linear Model Methodology*. Chapman & Hall/CRC, New York.

- Khuri, A., Mathew, T. and Sinha, B. (1998). *Statistical Tests for Mixed Linear Models* Wiley, New York.
- LaMotte, L. R. (1973). Quadratic Estimation of Variance Components. *Biometrics*. 29(2): 311–330.
- Lindstrom, M., Bates, D. (1988). Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data. *J. Am. Statist. Assoc.* 83: 1014–1022.
- Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Stat.* 5: 746–762.
- Pinheiro, J., Bates, D. (2000). *Mixed Effects Models in S and S-Plus*. Springer, New York.
- Searle, S. (1971). Topics in variance component estimation. *Biometrics*. 27: 1–76.
- Searle, S. (1995). An overview of variance component estimation. *Metrika*. 42: 215–230.
- Stern, S., Welsh, A. (2000). Likelihood Inference for Small Variance Components. *Can. J. Stat.* 28(3): 517–532.
- Ypma, T. (1995). Historical Development of the Newton-Raphson Method. *SIAM Review*. 37(4): 531–551.
- Vanleeuwen, D., Birkes, D., Seely, J. (1999). Balance and Orthogonality in Designs for Mixed Classification Models. *Ann. Stat.* 27(6): 1927–1947.