

A Dimension Reduction Technique for Estimation in Linear Mixed Models

M. de Carvalho[†], M. Fonseca[†], M. Oliveira[†], J.T. Mexia[†]

Abstract

This paper proposes a dimension reduction technique for estimation in linear mixed models. Specifically, we show that in a linear mixed model, the maximum likelihood problem can be rewritten as a substantially simpler optimization problem which presents at least two main advantages: the number of variables in the simplified problem is lower; the search domain of the simplified problem is a compact set. Whereas the former advantage reduces computational burden, the latter permits the use of stochastic optimization methods well qualified for closed bounded domains. The developed dimension reduction technique makes computation of maximum likelihood estimates, for fixed effects and variance components, feasible with large computational savings. Computational experience is reported here with the results evidencing an overall good performance of the proposed technique.

KEYWORDS: Maximum likelihood estimation; Linear mixed models; Stochastic optimization.

[†] Miguel de Carvalho (Email: mb.carvalho@fct.unl.pt) and Miguel Fonseca (Email: fmig@fct.unl.pt) are full researchers at CMA – Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal. Manuela Oliveira (Email: mmo@uevora.pt) is Professor at the Universidade de Évora. João Tiago Mexia (Email: jtm@fct.unl.pt) is Professor at the Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

1. INTRODUCTION

Maximum likelihood methods are among the main standard techniques for yielding parameter estimates for a statistical model of interest. The large sample characterization of this M -estimation methodology has long been established in the literature (Wald 1949). Despite the attractive features of maximum likelihood procedures, in a plurality of cases of practical interest, the proposed estimators are not analytically tractable. In order to overcome the lack of closed-form analytic solution, global optimization methods are typically employed. This feature is *not* peculiar to maximum likelihood estimation, as it is more generally shared by the broad class of extremum estimators, i.e., estimators which are motivated by an optimization problem of interest (Newey and McFadden 1994; Mexia and Corte-Real 2001). In this paper, we are concerned with a particular case where such an occurrence takes place, namely in maximum likelihood estimation of a linear mixed model. As we discuss below, this model is a popular extension of the linear model that is able to account for more than one source of error (see §2). A general overview of topics related with estimation and inference in linear mixed models, can be respectively found in Searle et al. (1992) and Khuri et al. (1998).

Before conducting estimation and inference, it is advisable to inspect if the problem at hand can be simplified analytically. Hence, for instance, as noted by Carvalho et al. (2007), if the linear mixed model has a common orthogonal block structure, then a closed form solution for the maximum likelihood estimator can in fact be found. Other than in these very special instances, it is seldom possible to achieve an explicit form for the solution of this maximum likelihood problem. Notwithstanding, here we show that in a linear mixed model, the maximum likelihood problem can be rewritten as a much simpler optimization problem (henceforth the simplified problem), whose size depends only on the number of variance components. The original maximum likelihood formulation is thus reduced into a simplified problem which presents at least two main advantages: the number of variables is considerably lower; the search domain is compact. As it can be readily appreciated, these features are extremely advantageous from the computational standpoint. In effect, this

simplified problem endows us with the means to obtain estimates of variance components with large computational savings. In addition, given that the search domain of the simplified problem is compact, we can rely the analysis over stochastic optimization methods well qualified for closed bounded domains (see, for instance, Spall 2003).

The remainder of this paper is structured as follows. Section 2 introduces the models of interest. Section 3 introduces the main result and presents simulation studies. Final remarks are provided in Section 4.

2. MODELS

2.1 The Linear Mixed Model

The simple linear model is at the heart of a broad number of statistical applications. In its elementary form, the model is commonly stated as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}, \quad (1)$$

where \mathbf{y} is the $(n \times 1)$ -vector of observations, \mathbf{X} is the design matrix of size $n \times k$, $\boldsymbol{\beta}_0$ is a $(k \times 1)$ -vector of unknown regression parameters, and $\boldsymbol{\epsilon}$ is $(n \times 1)$ -vector of unobserved errors. In this paper our interest relies in a well known generalization of the linear model (1) – namely, the linear mixed model. From the conceptual stance the model can be thought as an extension of the linear model that is qualified to consider distinct sources of error. Specifically, the linear mixed model takes the following form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \sum_{i=1}^{w-1} \mathbf{X}_i \boldsymbol{\zeta}_i + \boldsymbol{\epsilon}, \quad (2)$$

where $(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_0, \boldsymbol{\epsilon})$ are defined as in (1), \mathbf{X}_i are design matrices of size $n \times k_i$, and where $\boldsymbol{\zeta}_i$ are $(k_i \times 1)$ -vectors of unobserved random effects. Following classical assumptions, we take the random effects $\boldsymbol{\zeta}_i$ to be independent and normally distributed with null mean vectors and covariance matrix $\sigma_{0i}^2 \mathbf{I}_{k_i}$, for $i = 1, \dots, w-1$. Further, we also take the $\boldsymbol{\epsilon}$ to be normally distributed with null mean vector and covariance matrix $\sigma_{0w}^2 \mathbf{I}_n$, independently of the $\boldsymbol{\zeta}_i$, for

$i = 1, \dots, w - 1$. The model has the following mean vector and covariance matrix

$$E \{ \mathbf{y} | \mathbf{X} \} = \mathbf{X} \boldsymbol{\beta}_0$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\sigma}_0^2} = \sum_{i=1}^{w-1} \sigma_{0i}^2 \mathbf{X}_i \mathbf{X}_i' + \sigma_{0w}^2 \mathbf{I}_n,$$

where $\boldsymbol{\sigma}_0^2 = (\sigma_{01}^2, \dots, \sigma_{0w}^2)$. Given the current framework, we have that

$$\mathbf{y} | \mathbf{X} \sim \mathcal{N} \left(\mathbf{X} \boldsymbol{\beta}_0; \sum_{i=1}^{w-1} \sigma_{0i}^2 \mathbf{X}_i \mathbf{X}_i' + \sigma_{0w}^2 \mathbf{I}_n \right),$$

and thus the density for the model is given by

$$f_{\mathbf{y} | \mathbf{X}}(\mathbf{y}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_{\boldsymbol{\sigma}_0^2}^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_0)\right)}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_{\boldsymbol{\sigma}_0^2}|}}.$$

The estimator objective function assigned to the maximum likelihood estimator is given by the log-likelihood of the aforementioned linear mixed model, i.e.

$$\mathcal{L}_n(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\boldsymbol{\sigma}^2}| - \frac{1}{2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\sigma}^2}^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}).$$

The maximum likelihood estimators of the true parameter $\boldsymbol{\beta}_0$, and the model variance components $\boldsymbol{\sigma}^2$, respectively denoted by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}^2$, are thus given by

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\sigma}}^2 \end{pmatrix} = \arg \max_{(\boldsymbol{\beta}, \boldsymbol{\sigma}^2)' \in \mathbb{R}^k \times \mathbb{R}^w} \mathcal{L}_n(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \mathbf{y}). \quad (3)$$

In the following section, we consider a special case, wherein it is possible to obtain a closed form solution for this maximum likelihood problem. It should be emphasized that cases like the presented below represent the exception, rather than the rule. Notwithstanding, this instance is introduced here for serving as benchmark.

2.2 The Benchmark Case

In this subsection, we consider a particular case of model (2) wherein it is possible to get a closed form solution for the problem of interest. As we shall see below, this is gained at the

cost of the introduction of some structure in the covariance matrix $\Sigma_{\sigma_0^2}$. Specifically, if we consider the case wherein the covariance matrix can be decomposed as a linear combination of known orthogonal projection matrices \mathbf{Q}_j , i.e.

$$\Sigma_{\sigma_0^2} = \sum_{j=1}^w \eta_j \mathbf{Q}_j.$$

If the \mathbf{Q}_j are orthogonal projection matrices such that $\mathbf{Q}_j \mathbf{Q}_{j'} = \mathbf{0}$ for $j \neq j'$, and if \mathbf{T} , the orthogonal projection matrix on the range space of \mathbf{X} , is such that

$$\mathbf{T} \mathbf{Q}_j = \mathbf{Q}_j \mathbf{T}, \text{ for } j = 1, \dots, w,$$

the model is said to have commutative orthogonal block structure (Fonseca et al. 2006, 2010). Here, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_w)'$ is the vector of the so-called canonical variance components, and it is determined by the equation $\mathbf{B} \boldsymbol{\eta} = \boldsymbol{\sigma}_0^2$, where \mathbf{B} is a known nonsingular matrix. In this case, we can rewrite the density of the model as

$$f_{\mathbf{y}|\mathbf{X}}(\mathbf{y}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)' \left(\sum_{j=1}^w \eta_j^{-1} \mathbf{Q}_j\right) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)\right)}{\sqrt{(2\pi)^n \prod_{j=1}^w \eta_j^{g_j}}}.$$

where g_j is the rank of matrix \mathbf{Q}_j (Fonseca et al. 2006, Theorem 1). In this particular instance, it can be shown (Carvalho et al. 2007) that the estimators

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y},$$

and

$$\hat{\eta}_j = \frac{\mathbf{y}'(\mathbf{I} - \mathbf{T})\mathbf{Q}_j(\mathbf{I} - \mathbf{T})\mathbf{y}}{g_j},$$

solve the optimization problem (3).

When the above mentioned conditions on the covariance matrix do not hold, a closed-form analytical expression for producing maximum likelihood estimates is hardly available.

3. MAIN RESULT AND SIMULATION STUDY

3.1 Main Result

The next result establishes that in a linear mixed model, the maximum likelihood problem can be rewritten as a simplified problem where the search domain is a compact set whose dimension depends exclusively on the number of variance components. This result will be useful in order to compute the estimation of variance components, through maximum likelihood methods, with the application of random search methods.

Theorem 1 *The maximum likelihood estimators of the true parameter β_0 , and the model variance components σ_0^2 , respectively denoted by $\hat{\beta}$ and $\hat{\sigma}^2$*

$$\begin{pmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{pmatrix} = \arg \max_{(\beta, \sigma^2)' \in \mathbb{R}^k \times \mathbb{R}^w} \mathcal{L}_n(\beta, \sigma^2 | \mathbf{y}).$$

can be alternatively achieved by solving the following optimization problem

$$\min_{\gamma \in [0; \frac{\pi}{2}]^{w-1}} (f_n \circ \mathbf{p})(\gamma),$$

where

$$f_n(\alpha) = \ln \left(A(\alpha)^n |\Sigma_\alpha| \right) \tag{4}$$

$$A(\alpha) = \mathbf{y}' \left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\Sigma_\alpha^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_\alpha^{-1} \right)' \Sigma_\alpha^{-1} \left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\Sigma_\alpha^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_\alpha^{-1} \right) \mathbf{y} \tag{5}$$

$$\mathbf{p}(\gamma) = \mathbf{e}_1 \prod_{j=1}^{w-1} \cos(\gamma_j) + \sum_{l=2}^{w-1} \mathbf{e}_l \left(\prod_{j=1}^{w-l} \cos(\gamma_j) \sin(\gamma_{w-1}) \right) + \mathbf{e}_w \sin(\gamma_{w-1}), \tag{6}$$

and $\{\mathbf{e}_i\}_{i=1}^w$ denotes the canonical basis of \mathbb{R}^w .

As discussed above, we can ascertain at least two major advantages of the simplified problem, viz.:

- whereas the original maximum likelihood problem has dimension $(w+k)$, the simplified equivalent problem only has size $(w-1)$; additionally,

- the search domain of the simplified problem is $[0; \frac{\pi}{2}]^{w-1}$, and hence is compact (contrarily to what is verified in the original problem).

The proof is given in the appendix.

3.2 A First Monte Carlo Simulation Study

In this simulation study, we considered three one-way random models. (This study was implemented in R. The code is available from the authors.) The first model considered here is unbalanced with a total of 72 observations and 8 groups. The dissemination of the observations considered in the first model is described in Table 1. Several possible true values of the variance components were considered here.

Table 1: Description of the first one-way model considered (MODEL 1); the model is unbalanced with 72 observations segregated over 8 groups

Group	1	2	3	4	5	6	7	8
Observations	3	6	7	8	9	10	11	18

We then conducted a Monte Carlo simulation from which we report the averaging of the 1000 results achieved. In every run of the simulation, the optimization problem was solved by combining a random search technique with the dimension reduction technique introduced in §3.1. The random search algorithm used as input 10000 evaluations of the objective function of the simplified problem. Hence, 10000 values of $\gamma \in [0; \frac{\pi}{2}]$ were randomly selected and their corresponding images were generated via $(f_n \circ \mathbf{p})$.

Table 2: Estimates of the variance components in MODEL 1

Variance Component	0.00	0.10	0.50	0.70	1.00	1.50	2.00	5.00	10.00
Estimate	0.02	0.07	0.43	0.61	0.87	1.31	1.76	4.66	9.42

We now provide some guidelines to the interpretation of Table 2. In the first line we present the true values of the variance components σ_{01}^2 . The second line includes the solution provided by the recurrent application of random search methods to the optimization problem (3). Thus, for instance, when the “true” variance component was 0.5, the result yield through the application of stochastic optimization methods and the dimension reduction method is 0.425. Further, observe that except when the true value of the variance component is null, the true values always dominate the estimated values. Next, we considered a quasi-balanced model with 66 observations. The disposal of the observations was now given according to Table 3.

Table 3: Description of the second one-way model considered (MODEL II); the model is quasi-balanced with 66 observations segregated over 11 groups

Group	1	2	3	4	5	6	7	8	9	10	11
Observations	6	6	6	6	6	6	6	6	6	5	7

A Monte Carlo simulation was once more conducted. No changes were made to the true values of the variance components considered. The same applies to the methods used to perform the optimization step. The produced results are reported in Table 4.

Table 4: Estimates of the variance components in MODEL II

Variance Component	0.00	0.10	0.50	0.70	1.00	1.50	2.00	5.00	10.00
Estimate	0.02	0.09	0.45	0.63	0.89	1.34	1.85	4.61	9.09

Thus, for instance, when the true value of the variance component was 0.5, the estimate obtained was now 0.448. Again, it should be emphasized that, with the exception of the

case in which the true value of the variance component was 0, in all the remainder the true value was above the corresponding estimate.

A final model was then considered. The number of observations considered was now 72. Observations were grouped as described in Table 5.

Table 5: Description of the third one-way model considered (MODEL III); the model is unbalanced with 72 observations segregated over 9 groups

Group	1	2	3	4	5	6	7	8	9
Observations	2	2	3	3	4	4	15	15	24

The results are now summarized in Table 6.

Table 6: Estimates of the variance components in MODEL III

Variance Component	0.00	0.10	0.50	0.70	1.00	1.50	2.00	5.00	10.00
Estimate	0.01	0.07	0.43	0.60	0.85	1.36	1.78	4.71	9.93

As it can be readily noted, from the inspection of Tables 1, 2 and 3, there is a slight bias present in the estimates. This conforms with existing literature. In effect, there are some methods, such as Restricted Maximum Likelihood (REML), that can be used to compensate for such bias (see Harville 1974; Searle et al. 1992).

3.3 A Second Monte Carlo Simulation Study

In this section we report the results of a second Monte Carlo experiment. Here, a linear mixed model was considered with the following features: the fixed effects design matrix is of size ($n = 60 \times k = 9$); the random effects design matrix is of size ($n = 60 \times k_1 = 12$); both the fixed effects and the random effects matrices were randomly generated from a standard normal distribution. (This simulation study was implemented in Matlab. The routines

are available from the authors upon request.) This study provides a supplement for the application of the method based on Theorem 1.

In a similar way to the previous simulation study, the application of random search entailed the collection of 10000 sample points from the domain of the objective function $(f_n \circ \mathbf{p})$. Given the obvious increase in the burden of computation, here we considered 100 runs in this subsection. The true values of the variance components considered here were as follows: 0.1; 0.5; 0.75; 1; 2; 5. The results are summarized in Table 7. Here we also include the root mean square error (RMSE), the average absolute error (AAE) and the mean average percentage error (MAPE) of the estimates obtained in the several runs.

Table 7: Results from the second Monte Carlo simulation study

Variance Component	0.10	0.50	0.75	1.00	2.00	5.00
Estimate	0.10	0.51	0.77	0.98	1.93	4.84
RMSE	0.06	0.25	0.39	0.46	0.80	2.10
AAE	0.05	0.19	0.31	0.36	0.62	1.69

4. SUMMARY

A dimension reduction technique was proposed and applied in order to achieve, maximum likelihood estimates for the mixed model parameters and the variance components with large computational savings. The original maximum likelihood problem is thus reduced into a simplified problem which presents considerable computational advantages. In order to illustrate the mechanics of the proposed method two Monte Carlo simulations studies were here conducted.

APPENDIX

Proof of Theorem 1

Consider the log-likelihood of the aforementioned linear mixed model,

$$\mathcal{L}_n(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_{\sigma^2}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\sigma^2}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Observe that maximizing $\mathcal{L}_n(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \mathbf{y})$ is equivalent to minimizing

$$\mathcal{L}_*(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \mathbf{y}) = \ln |\boldsymbol{\Sigma}_{\sigma^2}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\sigma^2}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (7)$$

Now define $\boldsymbol{\sigma}^2 = c\boldsymbol{\alpha}$, with $c > 0$ and $\|\boldsymbol{\alpha}\| = 1$. Making use of the first-order conditions of the ML problem we get

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Sigma}_{\sigma^2}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{\sigma^2}^{-1} \mathbf{y}.$$

Hence, we can rewrite (7), evaluated at $\hat{\boldsymbol{\beta}}$, as

$$\mathcal{L}_* = n \ln(c) + \ln |\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}| + c^{-1} A(\boldsymbol{\alpha}),$$

where A is defined in (5). Now, observe that

$$\frac{\partial \mathcal{L}_*}{\partial c} = nc^{-1} - c^{-2} A(\boldsymbol{\alpha}) = 0 \Leftrightarrow c = \frac{A(\boldsymbol{\alpha})}{n},$$

and

$$\frac{\partial^2 \mathcal{L}_*}{\partial c^2} = 2c^{-3} A(\boldsymbol{\alpha}) - nc^{-2},$$

so that

$$\left. \frac{\partial^2 \mathcal{L}_*}{\partial c^2} \right|_{c=\frac{A(\boldsymbol{\alpha})}{n}} = \frac{n^3}{A(\boldsymbol{\alpha})} > 0,$$

whence

$$\hat{c} = \frac{A(\boldsymbol{\alpha})}{n},$$

is in fact an absolute minimum. Hence (7) simplifies into $n \ln(A(\boldsymbol{\alpha})) + \ln |\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|$, which we define as $f_n(\boldsymbol{\alpha})$ (see above in (4)). Next, we transform $\boldsymbol{\alpha}$ through the pseudo-polar coordinate transformation $\mathbf{p}(\boldsymbol{\gamma})$, as defined in (6); further details about this mapping can be found in Kendall (1961). This entails writing the w components of $\boldsymbol{\alpha}$, through $(w-1)$ components in $\boldsymbol{\gamma}$, as follows

$$\begin{aligned} \alpha_1 &= \cos(\gamma_1) \cdots \cos(\gamma_{w-2}) \cos(\gamma_{w-1}) \\ \alpha_2 &= \cos(\gamma_1) \cdots \cos(\gamma_{w-2}) \sin(\gamma_{w-1}) \\ &\vdots \\ \alpha_w &= \sin(\gamma_{w-1}). \end{aligned}$$

REFERENCES

- CARVALHO F, OLIVEIRA M and MEXIA, J (2007). “Maximum Likelihood Estimator in Models with Commutative Orthogonal Block Structure,” *Proceedings of 56th ISI Session*.
- CARVALHO F, MEXIA J and OLIVEIRA M (2009). “Estimation in Models with Commutative Orthogonal Block Structure,” *Journal of Statistical Theory and Practice*, 3, 523–533.
- FONSECA M, MEXIA J and ZMYŚLONY R (2006). “Binary Operations on Jordan algebras and Orthogonal Normal Models,” *Linear Algebra and Applications*, 417, 75–86.
- FONSECA M, MEXIA J and ZMYŚLONY R (2010). “Least Squares and Generalized Least Squares in Models with Orthogonal Block Structure,” *Journal of Statistical Planning and Inference*, 140, 1346–1352.
- HARVILLE D (1974). “Bayesian Inference for Variance Components Using Only Error Contrasts,” *Biometrika*, 61, 383–385.
- KENDALL M (1961). *A Course in the Geometry of n Dimensions*, Charles Griffin & Company Ltd.
- KHURI A, MATHEW T and SINHA B (1998). *Statistical Tests for linear mixed Models*, John Wiley & Sons.
- MEXIA J and CORTE-REAL P (2001). “Strong Law of Large Numbers for Additive Extremum Estimators,” *Discussiones Mathematicae - Probability and Statistics*, 21, 81–88.
- NEWBY W and MCFADDEN D (1994). “Large Sample Estimation and Hypothesis Testing,” *Handbook of Econometrics*, 4, 2111–2245.
- SEARLE S, CASELLA G and MCGULLOCH C (1992). *Variance Components*, John-Wiley and Sons, New York.
- SPALL J (2003). “Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control,” Wiley Interscience in Discrete Mathematics and Optimization.
- WALD A (1949). “Note on the Consistency of the Maximum Likelihood Estimate,” *Annals of Mathematical Statistics*, 20, 595–601.