# Linear mixed model implementation in lme4

Douglas Bates
Department of Statistics
University of Wisconsin – Madison
Bates@wisc.edu

January 27, 2006

**Abstract**

Expressions for the evaluation of the profiled log-likelihood or profiled log-restricted-likelihood of a linear mixed model, the gradients and Hessians of these criteria, and update steps for an ECME algorithm to optimize these criteria are given in Bates and DebRoy (2004). In this paper we generalize those formulae and describe the representation of mixed-effects models using sparse matrix methods available in the `Matrix` package.

## 1 Introduction

General formulae for the evaluation of the profiled log-likelihood and profiled log-restricted-likelihood in a linear mixed model are given in Bates and DebRoy (2004) and the use of a sparse matrix representation for such models is described in Bates (2004). The purpose of this paper is to describe the details of the implementation of this representation and those computational methods in the `lme4` package for R.

Because we concentrate on the computational methods and the representation, the order and style of presentation will be based on the sequence of calculations, not on the sequence in which the results would be derived. We will emphasize "what" and not "why". For the "why", refer to the papers cited above.

In §2 we describe the form and representation of the model. The calculation of the criteria to be optimized by the parameter estimates and related quantities is discussed in §**??**. Details of the calculation of the ECME step and the evaluation of the gradients of the criteria are given in §**??** and those of the Hessian in §**??**. In §**??** we give the details of an unconstrained parameterization for the model and the transformation of our results to this parameterization.

# 2 Form and representation of the model

We consider linear mixed models of the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}), \boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}^{-1}), \boldsymbol{\epsilon} \perp \boldsymbol{b} \tag{1}$$

where $\boldsymbol{y}$ is the $n$-dimensional response vector, $\boldsymbol{X}$ is an $n \times p$ model matrix for the $p$ dimensional fixed-effects vector $\boldsymbol{\beta}$, $\boldsymbol{Z}$ is the $n \times q$ model matrix for the $q$ dimensional random-effects vector $\boldsymbol{b}$ that has a Gaussian distribution with mean $\boldsymbol{0}$ and relative precision matrix $\boldsymbol{\Omega}$ (i.e., $\boldsymbol{\Omega}$ is the precision of $\boldsymbol{b}$ relative to the precision of $\boldsymbol{\epsilon}$), and $\boldsymbol{\epsilon}$ is the random noise assumed to have a spherical Gaussian distribution. The symbol $\perp$ indicates independence of random variables.

We will assume that $\boldsymbol{X}$ has full column rank and that $\boldsymbol{\Sigma}$ is positive definite.

## 2.1 Structure of the variance-covariance matrix

Components of the random effects vector $\boldsymbol{b}$ and portions of its variance-covariance matrix $\boldsymbol{\Sigma}$ are associated with $k$ grouping factors $\boldsymbol{f}_i, i = 1, \ldots, k$, each of length $n$, and with the $n_i, i = 1, \ldots, k$ levels of each of the grouping factors. In general there are $q_i$ components of $\boldsymbol{b}$ associated with each of the $n_i$ levels the grouping factor $\boldsymbol{f}_i, i = 1, \ldots, k$. Thus

$$q = \sum_{i=1}^{k} n_i q_i \tag{2}$$

We assume that the components of $\boldsymbol{b}$ and the rows and columns of $\boldsymbol{\Sigma}$ are ordered according to the $k$ grouping factors and, within the block for the $i$th grouping factor, according to the $n_i$ levels of the grouping factor.

Random effects associated with different grouping factors are independent. This implies that $\mathbf{\Sigma}$ is block-diagonal with $k$ diagonal blocks of orders $n_i q_i, i = 1, \ldots, k$.

Random effects associated with different levels of the same grouping factor are independent. This implies that the $i$th (outer) diagonal block of $\mathbf{\Sigma}$ is itself block diagonal in $n_i$ blocks of order $q_i$. We say that the structure of $\mathbf{\Sigma}$ is block/block diagonal.

Finally, the variance-covariance matrix within each of the $q_i$-dimensional subvectors of $\boldsymbol{b}$ associated with one of the $n_i$ levels of grouping factor $\boldsymbol{f}_i, i = 1, \ldots, k$ is a constant (but unknown) positive-definite symmetric $q_i \times q_i$ matrix $\mathbf{\Sigma}_i, i = 1, \ldots, k$. This implies that each of the $n_i$ inner diagonal blocks of order $q_i$ is a copy of $\mathbf{\Sigma}_i$. We say that $\mathbf{\Sigma}$ has a *repeated block/block diagonal* structure.

## 2.2 The relative precision matrix

Many of the computational formulae are more conveniently expressed in terms of $\mathbf{\Sigma}^{-1}$, which is called the *precision* matrix of the random effects, instead of $\mathbf{\Sigma}$. In fact, the formulae are most conveniently expressed in terms of the *relative precision matrix* $\sigma^2\mathbf{\Sigma}^{-1}$ which we write as $\mathbf{\Omega}$. That is,

$$\mathbf{\Omega} = \sigma^2\mathbf{\Sigma}^{-1} \tag{3}$$

This called the "relative" precision because it is precision of $\boldsymbol{b}$ ($\mathbf{\Sigma}^{-1}$) relative to the precision of $\boldsymbol{\epsilon}$ ($\boldsymbol{I}/(\sigma^2)$).

It is easy to establish that $\mathbf{\Omega}$ will have a repeated block/block diagonal structure like that of $\mathbf{\Sigma}$. That is, $\mathbf{\Omega}$ consists of $k$ outer diagonal blocks of sizes $n_i q_i, i = 1, \ldots, k$ and the $i$th outer diagonal block is itself block diagonal with $n_i$ inner blocks of size $q_i \times q_i$. Furthermore, each of the inner diagonal blocks in the $i$th outer block is a copy of the $q_i \times q_i$ positive-definite, symmetric matrix $\mathbf{\Omega}_i$.

Because $\mathbf{\Omega}$ has a repeated block/block structure we define the entire matrix if we specify the symmetric matrices $\mathbf{\Omega}_i, i = 1, \ldots, k$ and, because of the symmetry, $\mathbf{\Omega}_i$ has at most $q_i(q_i + 1)/2$ distinct elements. We will write $\boldsymbol{\theta}$ for a parameter vector of length at most $\sum_{i=1}^{k} q_i(q_i + 1)/2$ that determines $\mathbf{\Omega}$. For example, we could use the non-redundant elements in the $\mathbf{\Omega}_i$ as the components of $\boldsymbol{\theta}$. In fact we use a different, but equivalent, parameterization for reasons to be discussed later.

We only need to store the matrices $\mathbf{\Omega}_i, i = 1, \ldots, k$ and the number of levels in the grouping factors to be able create $\mathbf{\Omega}$. The matrices $\mathbf{\Omega}_i$ are stored in the `Omega` slot of an object of class `"lmer"`. The values of $k$ and $n_i, i = 1, \ldots, k$ can be determined from the list of the grouping factors themselves, stored as the `flist` slot, or from the dimensions $q_i, i = 1, \ldots, k$, stored in the `nc` slot, and the group pointers, stored in the `Gp` slot. Successive differences of the group pointers are the total number of components of $\boldsymbol{b}$ associated with the $i$th grouping factor. That is, these differences are $n_i q_i, i = 1, \ldots, k$.

## 2.3 Examples

Consider the fitted models

```
> Sm1 <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)
> data(Chem97, package = "mlmRev")
> Cm1 <- lmer(score ~ gcsescore + (1 | school) + (1 | lea), Chem97,
+     control = list(niterEM = 0, gradient = FALSE))
> data(star, package = "mlmRev")
> Mm1 <- lmer(math ~ gr + sx * eth + cltype + (yrs | id) + (1 |
+     tch) + (yrs | sch), star, control = list(niterEM = 0, gradient = FALSE))
```

Model `Sm1` has a single grouping factor with 18 levels. The `Omega` slot is a list of length one containing a $2 \times 2$ symmetric matrix.

```
> str(Sm1@flist)
List of 1
 $ Subject: Factor w/ 18 levels "308","309","310",..: 1 1 1 1 1 1 1 1 1 1 ...
> show(Sm1@Omega)
$Subject
2 x 2 Matrix of class "dpoMatrix"
            (Intercept)       Days
(Intercept)   1.0746247 -0.2942832
Days         -0.2942832 18.7549595
> show(Sm1@nc)
Subject
      2
> show(Sm1@Gp)
[1]  0 36
> diff(Sm1@Gp)/Sm1@nc
Subject
     18
```

Model `Cm1` has two grouping factors: the `school` factor with 2410 levels and the `lea` factor (local education authority - similar to a school district in the U.S.A.) with 131 levels. It happens that the `school` factor is nested within the `lea` factor, a property that we discuss below. The `Omega` slot is a list of length two containing two $1 \times 1$ symmetric matrices.

```
> str(Cm1@flist)
List of 2
 $ school: Factor w/ 2410 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ lea   : Factor w/ 131 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
> show(Cm1@Omega)
$school
1 x 1 Matrix of class "dpoMatrix"
            (Intercept)
(Intercept)    4.419673

$lea
1 x 1 Matrix of class "dpoMatrix"
            (Intercept)
(Intercept)    349.0502
> show(Cm1@nc)
school    lea
     1      1
> show(Cm1@Gp)
[1]    0 2410 2541
> diff(Cm1@Gp)/Cm1@nc
school    lea
  2410    131
```

Model `Mm1` has three grouping factor: `id` (the student) with 10732 levels, `tch` (the teacher) with 1374 levels and `sch` (the school) with 80 levels. The `Omega` slot is a list of length three containing two $2 \times 2$ symmetric matrices and one $1 \times 1$ matrix.

```
> str(Mm1@flist)
List of 3
 $ id : Factor w/ 10732 levels "100017","100028",..: 1 2 3 3 3 4 5 5 6 6 ...
 $ tch: Factor w/ 1374 levels "1","2","3","4",..: 476 889 695 698 703 1097 676 681 349 357 ...
 $ sch: Factor w/ 80 levels "1","2","3","4",..: 28 52 41 41 41 64 40 40 22 22 ...
> show(Mm1@Omega)
$id
2 x 2 Matrix of class "dpoMatrix"
            (Intercept)       yrs
(Intercept)   0.3320393 0.4956234
yrs           0.4956234 8.1873779

$tch
1 x 1 Matrix of class "dpoMatrix"
            (Intercept)
(Intercept)    1.425578

$sch
2 x 2 Matrix of class "dpoMatrix"
            (Intercept)       yrs
(Intercept)    3.289010  6.069119
yrs            6.069119 18.654882
> show(Mm1@nc)
```

```
 id tch sch
  2   1   2
> show(Mm1@Gp)
[1]     0 21464 22838 22998
> diff(Mm1@Gp)/Mm1@nc
   id   tch   sch
10732  1374    80
```

The last element of the `Gp` slot is the dimension of $\boldsymbol{b}$. Notice that for model `Mm1` the dimension of $\boldsymbol{b}$ is 22,998. This is also the order of the symmetric matrix $\boldsymbol{\Omega}$ but the contents of the matrix are determined by $\boldsymbol{\theta}$ which has a length of $3 + 1 + 3 = 7$ in this case.

# 3 Likelihood and restricted likelihood

# References


Douglas Bates. Sparse matrix representations of linear mixed models. *J. of Computational and Graphical Statistics*, 2004. submitted.

Douglas M. Bates and Saikat DebRoy. Linear mixed models and penalized least squares. *J. of Multivariate Analysis*, 2004. to appear.