

# A note on influence diagnostics in nonlinear mixed-effects elliptical models

Alexandre G. Patriota<sup>a,\*</sup>

<sup>a</sup>*Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo/SP, 05508-090, Brazil*

---

## Abstract

This paper provides general matrix formulas for computing the score function, the (expected and observed) Fisher information and the  $\Delta$  matrices (required for the assessment of local influence) for a quite general model which includes the one proposed by Russo *et al.* (2009). Additionally, we also present an expression for the generalized leverage on fixed and random effects. The matrix formulation has notational advantages, since although the complexity of the postulated model, all general formulas are compact, clear and have nice forms.

*Key words:* Elliptical models, Influence diagnostics, Matrix operations, Nonlinear models.

---

## 1. Main results

Recently, Russo *et al.* (2009) introduce an interesting nonlinear mixed model considering an elliptical distribution for the response variable. The authors also present a motivating example in a kinetics longitudinal data set which was firstly presented in Vonesh and Carter (1992) and previously analysed under the assumption of normality. Russo *et al.* (2009) analyse this dataset considering heavy-tailed distributions which may accommodate “large” observations. The authors compute the score function, Fisher information and some influence measures, but some matrices are presented only with the  $(r, s)$  element. Expressions for the entries of the expected Fisher information in a multivariate elliptical distribution was computed independently by Mitchell (1989) and Lange *et al.* (1989). Other recent papers have adopted the same strategy, namely Savalli *et al.* (2006) and Osorio *et al.* (2007). Since writing a matrix by entering element by element is not an efficient way to do it, we present a matrix version of these quantities (considering a more general model) in which, besides an aesthetic improvement, one can use it for avoiding that cumbersome task. Moreover, the compactness of the expressions might encourage other researches to study more complex models. We also show matrix versions of some expectations of a variable with an elliptical distribution that can be useful to apply in a multivariate context.

The nonlinear model studied in Russo *et al.* (2009) is given by

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\alpha}) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n \quad (1)$$

and, as defined by the authors,  $\mathbf{f}$  is an  $m_i$ -dimensional nonlinear function of  $\boldsymbol{\alpha}$ ,  $\mathbf{x}_i$  is a vector of covariates,  $\mathbf{Z}_i$  is a matrix of known constants,  $\boldsymbol{\alpha}$  is a  $p \times 1$  vector of unknown parameters and  $\mathbf{b}_i$  is an  $r \times 1$  vector of unobserved random regression coefficients, where  $(\mathbf{y}_i, \mathbf{b}_i)$  follows an elliptical distribution, such that

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{pmatrix} \stackrel{ind}{\sim} \text{El}_{m_i+r} \left[ \begin{pmatrix} \mathbf{f}(\mathbf{x}_i, \boldsymbol{\alpha}) \\ \mathbf{0} \end{pmatrix}; \begin{pmatrix} \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \sigma^2 \mathbf{I}_{m_i} & \mathbf{Z}_i \mathbf{D} \\ \mathbf{D} \mathbf{Z}_i^\top & \mathbf{D} \end{pmatrix} \right],$$

where  $\mathbf{I}_{m_i}$  is an  $(m_i \times m_i)$  identity matrix.

---

\*Corresponding author

*Email addresses:* patriota.alexandre@gmail.com and patriota@ime.usp.br (Alexandre G. Patriota)

For the purpose of avoiding numerical integrations, Russo *et al.* (2009) consider the marginal model, that is  $\mathbf{y}_i \stackrel{ind}{\sim} \text{El}_{m_i}(\mathbf{f}(\mathbf{x}_i, \boldsymbol{\alpha}); \boldsymbol{\Sigma}_i)$ , where  $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \sigma^2 \mathbf{I}_{m_i}$ . The vector of parameters of interest is defined as  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\gamma}^\top)^\top$ , where  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)^\top$  is the vector of parameters involved in  $\boldsymbol{\Sigma}_i$  with, in this case,  $\gamma_0 = \sigma^2$  and  $\gamma_1, \dots, \gamma_q$  are the parameters involved in  $\mathbf{D}$ . In addition to the authors' suppositions, the functional form of  $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\alpha})$  must be known and twice continuously differentiable with respect to each element of  $\boldsymbol{\alpha}$ .

In this paper, we consider the following model which was studied by Lange *et al.* (1989) considering a Student-t distribution,

$$\mathbf{y}_i \stackrel{ind}{\sim} \text{El}_{m_i}(\mathbf{f}(\mathbf{x}_i, \boldsymbol{\alpha}); \boldsymbol{\Sigma}_i(\mathbf{w}_i, \boldsymbol{\gamma})), \quad (2)$$

where  $\mathbf{w}_i$  and  $\mathbf{x}_i$  may have common components. The functional form of the covariance matrix  $\boldsymbol{\Sigma}_i(\mathbf{w}_i, \boldsymbol{\gamma})$  is known and twice continuously differentiable with respect to each element of  $\boldsymbol{\gamma}$ . Since  $\boldsymbol{\theta}$  must be identifiable in model (1), we suppose that the model fulfills this requirement. To see that model (1) is a special case of (2), take  $\mathbf{w}_i = \mathbf{Z}_i$  and  $\boldsymbol{\Sigma}_i(\mathbf{Z}_i, \boldsymbol{\gamma}) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \sigma^2 \mathbf{I}_{m_i}$ . As model (2) is not considering a specific structure for  $\boldsymbol{\Sigma}_i$ , it can represent other multivariate models. That is, model (1) can be generalized just by considering  $\mathbf{R}_i(\mathbf{z}_i, \boldsymbol{\sigma})$  instead of  $\sigma^2 \mathbf{I}_{m_i}$ , where  $\mathbf{z}_i$  is a vector of extra dispersion covariates. Then, in this context, we have that  $\boldsymbol{\Sigma}_i(\mathbf{w}_i, \boldsymbol{\gamma}) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \mathbf{R}_i(\mathbf{z}_i, \boldsymbol{\sigma})$  and  $\boldsymbol{\gamma} = (\boldsymbol{\tau}^\top, \boldsymbol{\sigma}^\top)^\top$ , where  $\mathbf{w}_i = (\mathbf{Z}_i^\top, \mathbf{z}_i^\top)^\top$ ,  $\boldsymbol{\tau}$  is a  $q_1 \times 1$  vector of dispersion parameters involved in  $\mathbf{D}$  and  $\boldsymbol{\sigma}$  is a  $q_2 \times 1$  vector of dispersion parameters associated with the model error term. We can go further and assign, for instance, a first-order autoregressive covariance matrix to the error terms, that is,  $\boldsymbol{\Sigma}_i(\mathbf{w}_i, \boldsymbol{\gamma}) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \sigma^2 \mathbf{V}(\rho)$ , where  $\mathbf{V}_{rs}(\rho) = \rho^{|r-s|}/(1-\rho)$ , then  $\mathbf{w}_i = \mathbf{Z}_i$ ,  $q_2 = 2$  and  $\boldsymbol{\gamma} = (\boldsymbol{\tau}, \sigma^2, \rho)^\top$ . The last structure is a common assumption when the vector  $\mathbf{y}_i$  has components measured in the time line, therefore the correlation of any pair of repeated measurements decreases according to how far apart they are in time. For instance, for  $r = q_1 = 1$  and  $m_i = m$  assume the following random effect model:  $y_{it} = \mu + \gamma_i + \epsilon_{it}$ , for  $i = 1, \dots, n$  and  $t = 1, \dots, m$ , where  $y_{it}$  is the measure made in the  $t$ th time for the  $i$ th individual. If  $t_1 < t_2 < \dots < t_m$ , we expect that  $\text{Cov}(y_{it_1}, y_{it_2}) > \text{Cov}(y_{it_1}, y_{it_3}) > \dots > \text{Cov}(y_{it_1}, y_{it_m})$ . Then, using the last structure for  $\boldsymbol{\Sigma}_i$  we obtain exactly this behavior, since in this case  $\text{Cov}(y_{it_1}, y_{it_s}) = \sigma_\gamma^2 + \sigma^2 \rho^{|t_1 - t_s|}/(1-\rho)$ . However, If we consider  $\mathbf{R}_i = \sigma^2 \mathbf{I}_m$ , that covariance will be  $\text{Cov}(y_{it_1}, y_{it_s}) = \sigma_\gamma^2 + \sigma^2$  for any pair of  $t_1$  and  $t_s$  not matter how far they are.

In general,  $\boldsymbol{\Sigma}_i(\mathbf{w}_i, \boldsymbol{\gamma})$  may be any structured covariance matrix with properties aforementioned. To keep the same notation, consider  $\boldsymbol{\gamma} = (\gamma_0, \dots, \gamma_q)^\top$ , i.e.,  $q_1 + q_2 = q + 1$ , then, the number of parameters is still  $b = p + q + 1$  (here,  $b$  is fixed and  $b \ll n$ ).

Russo *et al.* (2009) show that the score functions considering model (1) are given by

$$\mathbf{U}_\alpha = \sum_{i=1}^n v_i \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \quad \text{and} \quad \mathbf{U}_{\gamma_j} = -\frac{1}{2} \sum_{i=1}^n \left\{ \text{tr}[\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(j)}] - v_i \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(j)} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \right\} \quad \text{for } j = 0, \dots, q$$

where  $v_i = -2W_g(u_i)$ ,  $u_i = \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i$ ,  $\mathbf{r}_i = \mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \boldsymbol{\alpha})$ ,  $\mathbf{J}_i = \partial \mathbf{f}(\mathbf{x}_i, \boldsymbol{\alpha}) / \partial \boldsymbol{\alpha}^\top$ ,  $\dot{\boldsymbol{\Sigma}}_{i(j)} = \partial \boldsymbol{\Sigma}_i / \partial \gamma_j$ ,  $W_g(u_i) = d \log g(u_i) / du_i$  and function  $g(\cdot)$  is the density generator function with properties defined in Russo *et al.* (2009).

The authors also show that the expected Fisher information considering model (1) is given by

$$\mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{K}_{\alpha\alpha} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{\gamma\gamma} \end{pmatrix},$$

where

$$\mathbf{K}_{\alpha\alpha} = \sum_{i=1}^n \frac{4d_{gi}}{m_i} \mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{J}_i,$$

and the  $(r, s)$  element of  $\mathbf{K}_{\gamma\gamma}$  is given by

$$\mathbf{K}_{\gamma_r \gamma_s} = \sum_{i=1}^n \left\{ \frac{a_{rsi}}{4} (c_i - 1) + c_i \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(r)} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(s)}] \right\}$$

with  $c_i = 4f_{g_i}/[m_i(m_i + 2)]$  and the quantities  $d_{g_i}$ ,  $f_{g_i}$  and  $a_{r_{si}}$  are well defined in Russo *et al.* (2009). Note that, the above score functions and Fisher information are essentially the same of those under model (2), then the score functions and the Fisher information computed by Russo *et al.* (2009) hold even for more general models than the one that was defined by the authors.

This paper is organized as follows. Section 1.1 presents a matrix version for the score function, the (observed and expected) Fisher information and shows an iterative re-weighted least squares algorithm to attain the maximum-likelihood estimate for  $\boldsymbol{\theta}$ . Section 1.2 shows a matrix version for the  $\boldsymbol{\Delta}$  matrices presented by Russo *et al.* (2009) which are also applicable for model (2). Section 1.3 presents an expression for the generalized leverage on fixed effects considering model (2) and Section 1.4 shows a connection between the generalized leverage and the local influence. Section 1.5 discusses the generalized leverage on random effects. Additionally, Section 2 applies the generalized leverage on fixed and random effects for the same dataset analysed by Russo *et al.* (2009) and Section 3 conducts a small simulation study to verify the computational speed of the matrix and element-by-element formulations. It is worth emphasizing that this work should be seen just as a complementary material of Russo *et al.* (2009).

### 1.1. Matrix version for the score function and Fisher information

The following two matrix results will be intensively used in the computation of the expressions derived in this paper. Let  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  be  $n \times n$  matrices, define also  $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  and  $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n)$ , where  $\mathbf{a}_i$  and  $\mathbf{c}_i$  are  $n \times 1$  vectors, then

$$\text{tr}\{\mathbf{A}^\top \mathbf{C} \mathbf{D} \mathbf{B}^\top\} = \text{vec}(\mathbf{A})^\top (\mathbf{B} \otimes \mathbf{C}) \text{vec}(\mathbf{D}) \quad \text{and} \quad \mathbf{A}^\top \mathbf{B} \mathbf{C} = \{\mathbf{a}_r^\top \mathbf{B} \mathbf{c}_s\} \quad (3)$$

where  $\text{vec}(\cdot)$  is the vec operator, which transforms a matrix into a vector by stacking the columns of the matrix one underneath the other, “ $\otimes$ ” indicates the Kronecker product. These results and other methods in matrix differential calculus can be studied in Magnus and Neudecker (2007).

Define the following quantities,

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{J}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_i \end{pmatrix}, \quad \mathbf{H}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \boldsymbol{\Sigma}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1} \end{pmatrix}, \quad \dot{\mathbf{u}}_i = \begin{pmatrix} v_i \mathbf{r}_i \\ -\text{vec}(\boldsymbol{\Sigma}_i - v_i \mathbf{r}_i \mathbf{r}_i^\top) \end{pmatrix}$$

and  $\mathbf{V}_i = (\text{vec}(\dot{\boldsymbol{\Sigma}}_{i(0)}), \dots, \text{vec}(\dot{\boldsymbol{\Sigma}}_{i(q)}))$ , where  $(\mathbf{F}_1^\top, \dots, \mathbf{F}_n^\top)^\top$  has rank  $b$  (i.e., the functions  $\mathbf{f}$  and  $\boldsymbol{\Sigma}_i$  must be defined to hold such condition). Then, by using (3) and after a somewhat algebra, we have that the score function and the expected Fisher information, considering model (2), can be written, respectively, as

$$\mathbf{U}_\theta = \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \dot{\mathbf{u}}_i \quad \text{and} \quad \mathbf{K}_{\theta\theta} = \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{H}_i \mathbf{O}_i \mathbf{H}_i \mathbf{F}_i \quad (4)$$

where

$$\mathbf{O}_i = c_i \begin{pmatrix} \frac{4d_{g_i}}{m_i c_i} \boldsymbol{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & 2\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i \end{pmatrix} + (c_i - 1) \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \text{vec}(\boldsymbol{\Sigma}_i) \text{vec}(\boldsymbol{\Sigma}_i)^\top \end{pmatrix}.$$

Fisher information given in (4) can clearly be interpreted as a quadratic form which can be easily attained through direct matrix operations. Thus, a joint iterative procedure for attaining the MLE of  $\boldsymbol{\theta}$  can be formulated as the following re-weighted least squares algorithm

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \left( \sum_{i=1}^n \mathbf{F}_i^{(m)\top} \widetilde{\mathbf{H}}_i^{(m)} \mathbf{F}_i^{(m)} \right)^{-1} \left( \sum_{i=1}^n \mathbf{F}_i^{(m)\top} \widetilde{\mathbf{H}}_i^{(m)} \widetilde{\mathbf{u}}_i^{(m)} \right), \quad m = 1, 2, \dots \quad (5)$$

where the quantities with the upper script “ $(m)$ ” are evaluated at  $\hat{\boldsymbol{\theta}}^{(m)}$ ,  $\widetilde{\mathbf{H}}_i = \mathbf{H}_i \mathbf{O}_i \mathbf{H}_i$ ,  $\widetilde{\mathbf{u}}_i = \mathbf{H}_i^{-1} \mathbf{O}_i^{-1} \dot{\mathbf{u}}_i + \mathbf{F}_i \hat{\boldsymbol{\theta}}$  and  $m$  is the iteration counter. Under normality we have that  $c_i = 1$ ,  $\mathbf{O}_i = \mathbf{H}_i^{-1}$  and  $v_i = 1$ , and it is easy to see that this iterative procedure (under normality) is a special case of the one proposed in Patriota and Lemonte (2009).

In the sequence, we provide a matrix formulation for the observed Fisher information which requires harder matrix operations than the one spent in the expected Fisher information. The observed Fisher information presented in Russo *et al.* (2009), that is the same observed Fisher information considering model (2), is given by  $-\ddot{\mathbf{L}}_{\theta\theta} = -\sum_{i=1}^n \ddot{\mathbf{L}}_{\theta\theta,i}$ , with

$$\ddot{\mathbf{L}}_{\theta\theta,i} = \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{\mathbf{L}}_{\alpha\alpha,i} & \ddot{\mathbf{L}}_{\alpha\gamma,i} \\ \ddot{\mathbf{L}}_{\gamma\alpha,i} & \ddot{\mathbf{L}}_{\gamma\gamma,i} \end{pmatrix}$$

where

$$\begin{aligned} \ddot{\mathbf{L}}_{\alpha\alpha,i} &= 2\mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \{W_g(u_i)\boldsymbol{\Sigma}_i + 2W'_g(u_i)\mathbf{r}_i\mathbf{r}_i^\top\} \boldsymbol{\Sigma}_i^{-1} \mathbf{J}_i - 2W_g(u_i)[\mathbf{I}_p \otimes \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1}] \mathbf{D}_i, \\ \ddot{\mathbf{L}}_{\alpha\gamma,i} &= (\ddot{\mathbf{L}}_{\alpha\gamma_0,i}, \ddot{\mathbf{L}}_{\alpha\gamma_1,i}, \dots, \ddot{\mathbf{L}}_{\alpha\gamma_q,i}) \end{aligned} \quad (6)$$

with  $\ddot{\mathbf{L}}_{\alpha\gamma_j,i} = 2\mathbf{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \{W_g(u_i)\boldsymbol{\Sigma}_i + W'_g(u_i)\mathbf{r}_i\mathbf{r}_i^\top\} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(j)} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i$  and the element  $(j, k)$  of  $\ddot{\mathbf{L}}_{\gamma\gamma,i}$  has the form

$$\begin{aligned} \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}_i^{-1} \left[ \dot{\boldsymbol{\Sigma}}_{i(j)} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(k)} - \dot{\boldsymbol{\Sigma}}_{i(jk)} \right] \right\} + \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \left\{ W'_g(u_i) \dot{\boldsymbol{\Sigma}}_{i(j)} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(k)} - W_g(u_i) \dot{\boldsymbol{\Sigma}}_{i(jk)} \right. \\ \left. + W_g(u_i) \dot{\boldsymbol{\Sigma}}_{i(j)} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(k)} + W_g(u_i) \dot{\boldsymbol{\Sigma}}_{i(k)} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(j)} \right\} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \end{aligned} \quad (7)$$

with

$$\dot{\boldsymbol{\Sigma}}_{i(jk)} = \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial \gamma_j \partial \gamma_k}, \quad \mathbf{D}_i = \begin{pmatrix} \mathbf{a}_{i(11)} & \dots & \mathbf{a}_{i(1p)} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{i(p1)} & \dots & \mathbf{a}_{i(pp)} \end{pmatrix} \quad \text{and} \quad \mathbf{a}_{i(rs)} = \frac{\partial^2 \mathbf{f}}{\partial \alpha_r \partial \alpha_s}.$$

Note that, quantities (6) and (7) are not written in a matrix form, in the following we present a compact matrix version of  $\ddot{\mathbf{L}}_{\theta\theta}$ .

$$\ddot{\mathbf{L}}_{\theta\theta} = \sum_{i=1}^n \left\{ \mathbf{F}_i^\top \mathbf{H}_i \ddot{\mathbf{O}}_i \mathbf{H}_i \mathbf{F}_i + [\dot{\mathbf{u}}_i^\top \mathbf{H}_i] \left[ \frac{\partial \mathbf{F}_i}{\partial \boldsymbol{\theta}} \right] \right\} \quad (8)$$

where

$$\begin{aligned} \ddot{\mathbf{O}}_i &= 2W_g(u_i) \begin{pmatrix} \boldsymbol{\Sigma}_i & & \\ 2\boldsymbol{\Sigma}_i \otimes \mathbf{r}_i & 2\boldsymbol{\Sigma}_i \otimes \mathbf{r}_i^\top & \\ & (\mathbf{r}_i \mathbf{r}_i^\top) \otimes \boldsymbol{\Sigma}_i & \end{pmatrix} + 2 \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i \end{pmatrix} \\ &+ 4W'_g(u_i) \begin{pmatrix} \mathbf{r}_i \mathbf{r}_i^\top & & \\ (\mathbf{r}_i \mathbf{r}_i^\top) \otimes \mathbf{r}_i & \text{vec}(\mathbf{r}_i \mathbf{r}_i^\top) \otimes \mathbf{r}_i^\top & \\ & \text{vec}(\mathbf{r}_i \mathbf{r}_i^\top) \text{vec}(\mathbf{r}_i \mathbf{r}_i^\top)^\top & \end{pmatrix}, \end{aligned}$$

$\frac{\partial \mathbf{F}_i}{\partial \boldsymbol{\theta}}$  is an  $m_i(m_i + 1) \times b \times b$  array,  $[\dot{\mathbf{u}}_i^\top \mathbf{H}_i] \left[ \frac{\partial \mathbf{F}_i}{\partial \boldsymbol{\theta}} \right]$  is the bracket product of  $\dot{\mathbf{u}}_i^\top \mathbf{H}_i$  and  $\frac{\partial \mathbf{F}_i}{\partial \boldsymbol{\theta}}$  (for further details see Wei, 1998, on pg. 188).

In what follows, we present some matrix results on elliptical variables. Here,  $\mathbf{r}_i \stackrel{\text{ind}}{\sim} \text{El}_{m_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ , then adapting the results of Mitchell (1989) for a matrix version, we have that

- a)  $E(\mathbf{r}_i v_i) = \mathbf{0}$ ,
- b)  $E(\mathbf{r}_i \mathbf{r}_i^\top v_i) = \boldsymbol{\Sigma}_i$ ,
- c)  $E(\mathbf{r}_i \mathbf{r}_i^\top v_i^2) = 4d_{gi}/m_i \boldsymbol{\Sigma}_i$
- d)  $E(\text{vec}(\mathbf{r}_i \mathbf{r}_i^\top) \mathbf{r}_i^\top v_i^2) = \mathbf{0}$
- e)  $E(\text{vec}(\mathbf{r}_i \mathbf{r}_i^\top) \text{vec}(\mathbf{r}_i \mathbf{r}_i^\top)^\top v_i^2) = c_i \left( \text{vec}(\boldsymbol{\Sigma}_i) \text{vec}(\boldsymbol{\Sigma}_i)^\top + \boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i + \mathbf{P}_i(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i) \right)$ , where  $\mathbf{P}_i$  is a comutation maltrix such that  $\text{vec}(\mathbf{A}) = \mathbf{P}_i \text{vec}(\mathbf{A}^\top)$  for any matrix  $\mathbf{A}$  with appropriated dimensions.

Therefore, as we are considering a function  $g(\cdot)$  with regular properties (differentiation and integration are interchangeable), we have that  $E(\dot{\mathbf{u}}_i) = \mathbf{0}$ , and  $E(-\ddot{\mathbf{L}}_{\theta\theta}) = \mathbf{K}_{\theta\theta}$ .

### 1.2. Matrix version for $\Delta$

The diagnostic technique developed in Cook (1986) is a well-spread tool to check the model assumptions and conduct diagnostic studies. The author proposes to look at the likelihood displacement  $LD(\boldsymbol{\omega}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$  to find possible influential observations in the MLEs, where  $L(\boldsymbol{\theta}) = \sum_i L_i(\boldsymbol{\theta})$  is the log-likelihood function and  $\boldsymbol{\omega}$  is an  $s \times 1$  vector of perturbation restricted in an open set  $\Omega \subset \mathbb{R}^s$ . It is also defined a vector of no perturbation as  $\boldsymbol{\omega}_0 \in \Omega$  in which  $LD(\boldsymbol{\omega}_0) = 0$ , i.e.,  $L(\boldsymbol{\theta}_{\boldsymbol{\omega}_0}) = L(\boldsymbol{\theta})$ . In his seminal paper, Cook shows that the normal curvature at the unit direction  $\boldsymbol{\ell}$  has the following form  $C_{\boldsymbol{\ell}}(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^\top \boldsymbol{\Delta}^\top (\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}})^{-1} \boldsymbol{\Delta} \boldsymbol{\ell}|$  where  $\boldsymbol{\Delta} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$ , both  $\boldsymbol{\Delta}$  and  $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$  are evaluated at  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  and  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ . Thus,  $C_{\mathbf{d}_{\max}}$  is twice the largest eigenvalue of  $\mathbf{B} = -\boldsymbol{\Delta}^\top \ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \boldsymbol{\Delta}$  and  $\mathbf{d}_{\max}$  is the corresponding eigenvector. The index plot of  $\mathbf{d}_{\max}$  may reveal how to perturb the model (or data) to obtain large changes in the estimate of  $\boldsymbol{\theta}$ . For a more detailed information, we refer the reader to the work of Russo *et al.* (2009) and the references therein.

The main goal of this section is to give matrix versions of the  $\boldsymbol{\Delta}$  matrices computed by Russo *et al.* (2009) for each perturbation scheme. Note that, by using the defined quantities, we can write the  $b \times n$  matrix  $\boldsymbol{\Delta}$  in the case-weight perturbation (i.e.,  $L_i(\boldsymbol{\theta}_{\boldsymbol{\omega}_0}) = \omega_i L_i(\boldsymbol{\theta})$ ) and the scale perturbation (i.e., the perturbed log-likelihood function  $L_i(\boldsymbol{\theta}_{\boldsymbol{\omega}})$  is built replacing  $\boldsymbol{\Sigma}_i$  with  $\omega_i^{-1} \boldsymbol{\Sigma}_i$  in  $L_i(\boldsymbol{\theta})$ ), respectively, by

$$\boldsymbol{\Delta} = \left( \hat{\mathbf{F}}_1^\top \hat{\mathbf{H}}_1 \hat{\mathbf{u}}_1, \dots, \hat{\mathbf{F}}_n^\top \hat{\mathbf{H}}_n \hat{\mathbf{u}}_n \right) \quad \text{and} \quad \boldsymbol{\Delta} = \left( \hat{\mathbf{F}}_1^\top \hat{\mathbf{H}}_1 \hat{\mathbf{v}}_1, \dots, \hat{\mathbf{F}}_n^\top \hat{\mathbf{H}}_n \hat{\mathbf{v}}_n \right), \quad (9)$$

where the quantities with “ $\hat{\phantom{x}}$ ” are evaluated at  $\hat{\boldsymbol{\theta}}$  and

$$\hat{\mathbf{v}}_i = -2(W_g(u_i) + u_i W'_g(u_i)) \begin{pmatrix} \mathbf{r}_i \\ \text{vec}(\mathbf{r}_i \mathbf{r}_i^\top) \end{pmatrix}.$$

Finally, the  $b \times N$  matrix  $\boldsymbol{\Delta}$  under the response perturbation (i.e., the perturbed log-likelihood function  $L_i(\boldsymbol{\theta}_{\boldsymbol{\omega}})$  is built replacing  $\mathbf{y}_i$  with  $\mathbf{y}_i + \boldsymbol{\omega}_i$  in  $L_i(\boldsymbol{\theta})$ ) becomes

$$\boldsymbol{\Delta} = \left( \hat{\mathbf{F}}_1^\top \hat{\mathbf{H}}_1 \hat{\mathbf{G}}_1, \dots, \hat{\mathbf{F}}_n^\top \hat{\mathbf{H}}_n \hat{\mathbf{G}}_n \right), \quad (10)$$

where  $N = \sum_{i=1}^n m_i$  and

$$\hat{\mathbf{G}}_i = -2 \begin{pmatrix} W_g(u_i) \mathbf{I}_{m_i} + 2W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \\ 2\mathbf{r}_i \otimes (W_g(u_i) \mathbf{I}_{m_i} + W'_g(u_i) \mathbf{r}_i \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1}) \end{pmatrix}.$$

Note that, formulas (9) and (10) are easily handled through any statistical software.

### 1.3. Generalized leverage on fixed effects

In this section, we compute the generalized leverage proposed by Wei *et al.* (1998). The generalized leverage is a measure of the importance of individual observations on the estimator. It is one of the measures of influence analysis in regression models. Let  $\mathbf{y} = \text{vec}(\mathbf{y}_1, \dots, \mathbf{y}_n)$  and  $\boldsymbol{\mu}(\boldsymbol{\alpha}) = \text{vec}(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_1), \dots, \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_n))$ . The authors have shown that the generalized leverage is obtained by evaluating the  $N \times N$  matrix

$$\mathbf{GL}(\boldsymbol{\theta}) = \mathbf{D}_{\boldsymbol{\theta}} (-\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}})^{-1} \ddot{\mathbf{L}}_{\boldsymbol{\theta}\mathbf{y}},$$

at  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , where  $\mathbf{D}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}(\boldsymbol{\alpha})/\partial\boldsymbol{\theta}^\top$  and  $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\mathbf{y}} = \partial^2 L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\mathbf{y}^\top$ . The main idea behind the concept of leverage is that of evaluating the influence of  $\mathbf{y}_i$  on its own predicted value. As noted by the authors, the generalized leverage is invariant under reparameterizations and observations with large  $GL_{ii}$  are leverage points. Under the model defined in (2), we have that

$$\mathbf{D}_{\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} \\ \mathbf{J}_2 & \mathbf{0} \\ \vdots & \vdots \\ \mathbf{J}_n & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \ddot{\mathbf{L}}_{\boldsymbol{\theta}\mathbf{Y}} = \left( \mathbf{F}_1^\top \mathbf{H}_1 \mathbf{G}_1, \dots, \mathbf{F}_n^\top \mathbf{H}_n \mathbf{G}_n \right).$$

Consider now that  $\gamma$  is fixed. Then, the generalized leverage is

$$GL(\alpha) = D_\alpha(-\ddot{L}_{\alpha\alpha})^{-1}\ddot{L}_{\alpha y}, \quad (11)$$

at  $\alpha = \hat{\alpha}$ , where  $D_\alpha = \partial\mu(\alpha)/\partial\alpha^\top$  and  $\ddot{L}_{\alpha y} = \partial^2 L(\alpha)/\partial\alpha\partial y^\top$ . Under model (2), we find  $D_\alpha = (\mathbf{J}_1^\top, \dots, \mathbf{J}_n^\top)^\top$  and  $\ddot{L}_{\alpha y} = (\mathbf{J}_1^\top \Sigma_1^{-1} \mathbf{G}_{11}, \dots, \mathbf{J}_n^\top \Sigma_n^{-1} \mathbf{G}_{1n})$ , where  $\mathbf{G}_{1i} = -2(W_g(u_i)\mathbf{I}_{m_i} + 2W'_g(u_i)\mathbf{r}_i\mathbf{r}_i^\top \Sigma_i^{-1})$ . Index plots of  $GL_{ii}$  may reveal those observations with high influence on their own predicted values.

#### 1.4. Connection between local influence and generalized leverage

There is a connection between local influence and generalized leverage. In order to deduce such relationship we must define some quantities. Define  $\mathbf{F}_i = (\mathbf{F}_{1i}^\top, \mathbf{F}_{2i}^\top)^\top$ ,  $\mathbf{G}_i = (\mathbf{G}_{1i}^\top, \mathbf{G}_{2i}^\top)^\top$  where  $\mathbf{F}_{1i} = (\mathbf{J}_i, \mathbf{0})$ ,  $\mathbf{F}_{2i} = (\mathbf{0}, \mathbf{V}_i)$ ,  $\mathbf{G}_{1i} = -2(W_g(u_i)\mathbf{I}_{m_i} + 2W'_g(u_i)\mathbf{r}_i\mathbf{r}_i^\top \Sigma_i^{-1})$  and  $\mathbf{G}_{2i} = -2[2\mathbf{r}_i \otimes (W_g(u_i)\mathbf{I}_{m_i} + W'_g(u_i)\mathbf{r}_i\mathbf{r}_i^\top \Sigma_i^{-1})]$ . Define also,

$$\mathbf{F}^* = (\mathbf{F}_1^\top, \dots, \mathbf{F}_n^\top)^\top \quad \text{and} \quad \tilde{\mathbf{F}} = (\tilde{\mathbf{F}}_1^\top, \tilde{\mathbf{F}}_2^\top)^\top,$$

where  $\tilde{\mathbf{F}}_j = (\mathbf{F}_{j1}^\top, \dots, \mathbf{F}_{jn}^\top)^\top$  for  $j = 1, 2$ . Notice that, there exist a matrix of permutations  $\mathbf{I}^*$  such that  $\mathbf{F}^* = \mathbf{I}^* \tilde{\mathbf{F}}$ . This matrix is given by

$$\mathbf{I}^* = \begin{pmatrix} \mathbf{I}_{m_1} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{I}_{m_1^2} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{m_2^2} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{m_{n-1}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_{m_{n-1}^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_{m_n} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_{m_n^2} \end{pmatrix},$$

where the matrices of zeros have appropriated dimensions. Therefore, we have that

$$\Delta^\top(-\ddot{L}_{\theta\theta})^{-1}\Delta = \mathbf{P}^\top \mathbf{I}^* \tilde{\mathbf{F}}(-\ddot{L}_{\theta\theta})^{-1}\Delta,$$

where  $\mathbf{P} = \text{blc diag}(\mathbf{H}_1 \mathbf{G}_1, \dots, \mathbf{H}_n \mathbf{G}_n)$ . Note also that,  $\mathbf{P}^\top \mathbf{I}^* = (\mathbf{P}_1^{*\top}, \mathbf{P}_2^{*\top})$ , where

$$\mathbf{P}_1^* = \text{blc diag}(\Sigma_1^{-1} \mathbf{G}_{11}, \dots, \Sigma_n^{-1} \mathbf{G}_{1n}) \quad \text{and} \quad \mathbf{P}_2^* = 1/2 \text{blc diag}((\Sigma_1^{-1} \otimes \Sigma_1^{-1}) \mathbf{G}_{21}, \dots, (\Sigma_n^{-1} \otimes \Sigma_n^{-1}) \mathbf{G}_{2n}).$$

Hence, the relationship between the local influence (under additive response perturbations) and the generalized leverage is

$$\Delta^\top(-\ddot{L}_{\theta\theta})^{-1}\Delta = \mathbf{P}_1^{*\top} \mathbf{GL}(\theta) + \mathbf{P}_2^{*\top} \tilde{\mathbf{F}}_2(-\ddot{L}_{\theta\theta})^{-1}\Delta.$$

This connection was studied by Osorio *et al.* (2007) for the linear case considering fixed the dispersion parameters. Thus, following the same idea we can consider  $\gamma$  fixed and find the generalized leverage for  $\alpha$ . The matrix  $\mathbf{B}$  of the local influence under the response perturbation considering  $\gamma$  fixed is  $\mathbf{B} = \Delta_\alpha^\top(-\ddot{L}_{\alpha\alpha})^{-1}\Delta_\alpha$ , where  $\Delta_\alpha = \ddot{L}_{\alpha y}$  with  $\ddot{L}_{\alpha y}$  defined in the previous section. Thus, the connection between the local influence and the generalized leverage (for  $\gamma$  fixed) is  $\mathbf{B} = \mathbf{P}_1^{*\top} \mathbf{GL}(\alpha)$ .

Then, the normal curvature under additive perturbations in the response values can be rewritten as  $C_\ell(\theta) = 2|\ell^\top \mathbf{P}_1^{*\top} \mathbf{GL}(\theta)\ell - \ell^\top \mathbf{P}_2^{*\top} \tilde{\mathbf{F}}_2(-\ddot{L}_{\theta\theta})^{-1}\Delta\ell|$  in the general case and  $C_\ell(\alpha) = 2|\ell^\top \mathbf{P}_1^{*\top} \mathbf{GL}(\alpha)\ell|$  for  $\gamma$  fixed.

### 1.5. Generalized leverage on random effects

In this section we apply the idea of Nobre and Singer (2010) to verify the importance of individual observations on the random effect components of the conditional fitted values  $\tilde{\mathbf{y}} = \boldsymbol{\mu}(\hat{\boldsymbol{\alpha}}) + \mathbf{Z}\hat{\mathbf{b}}$  for  $\boldsymbol{\gamma}$  fixed, where  $\mathbf{Z} = \text{blc diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$  and  $\hat{\mathbf{b}} = \text{vec}(\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n)$  with the predictor  $\hat{\mathbf{b}}_i = \mathbf{D}\mathbf{Z}_i^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i, \hat{\boldsymbol{\alpha}}))$  computed by Russo *et al.* (2009). Therefore, in matrix notation we have

$$\mathbf{GL}_p = \frac{\partial \tilde{\mathbf{y}}}{\partial \mathbf{y}^\top} = \mathbf{GL}(\boldsymbol{\alpha}) + \mathbf{GL}_b$$

where  $\mathbf{GL}_b = \mathbf{Z}(\mathbf{I}_n \otimes \mathbf{D})\mathbf{Z}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{I}_N - \mathbf{GL}(\boldsymbol{\alpha}))$ ,  $\boldsymbol{\Sigma} = \text{blc diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n)$  and  $\mathbf{GL}(\boldsymbol{\alpha})$  given in (11). The first idea for the generalized leverage on random effect effects was proposed independently by Nobre (2004) and Demidenko and Stukel (2005). They suggest using the quantity  $\mathbf{GL}_b$ . However, as noted by Nobre and Singer (2010) this quantity carries the leverages for the fixed effects, therefore the same authors suggest using the quantity  $\mathbf{GL}_{b^*} = \mathbf{Z}(\mathbf{I}_n \otimes \mathbf{D})\mathbf{Z}^\top$  rather than  $\mathbf{GL}_b$  to measure the leverage on random effects. It is very reasonable, since outlying observations in the vector space spanned by the explanatory variables in  $\mathbf{Z}$  should affect the random effect component of the conditional fitted values and consequently affect the estimate of the within-unit variability explained by the presence of the random effects (Nobre and Singer, 2010). Then the  $i$ th individual have high leverage on the random effects if  $\text{tr}\{\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top\}/m_i$  is sufficiently large. The  $j$ th observation of the  $i$ th individual is said to have high leverage on the random effects if the  $j$ th diagonal element of  $\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top$  is sufficiently large. Sufficiently large can be understood as greater than  $2\text{tr}\{\mathbf{GL}_{b^*}/N\}$ .

## 2. Application of the Generalized leverage

This section applies the generalized leverage on the dataset analysed by Russo *et al.* (2009). This dataset is called ‘Dialyzer’ and can be found in the library ‘nlme’ of the software R Development Core Team (2009). As expected, the maximum likelihood estimates and their standard deviations obtained by using the matrix formulation are the very same of those given in Table 1 of Russo *et al.* (2009).

Figure 1 presents index plots of leverages on the fixed and random effects, where (a), (b) and (c) refer to the fixed effects considering Normal, Student-t and Power-Exponential distributions, respectively and (d), (e) and (f) refer to the random effects considering Normal, Student-t and Power-Exponential distributions, respectively. As cut-off points we consider  $2\text{tr}\{\mathbf{GL}(\boldsymbol{\theta})/N\}$  for the fixed effects and  $2\text{tr}\{\mathbf{GL}_{b^*}/N\}$  for the random effects (the dashed line in the Figure 1), we also consider  $Q_3 + 1.5(Q_3 - Q_1)$  (the full line in Figure 1), where  $Q_3$  is the third quartile and  $Q_1$  is the first quartile of the diagonal elements of  $\mathbf{GL}(\boldsymbol{\theta})$  for the fixed effects and  $\mathbf{GL}_{b^*}$  for the random effects. The latter cut-off point is used in the boxplot graph to show the outliers of a dataset and it takes into account the dispersion of the points. As can be seen, there is no evident leverages on the random effects. For the leverages on the fixed effects and considering the cut-off point based on the quartiles, we find that 13 points seem to be leverages under the Normal distribution (1.1, 2.1, 3.1, 4.1, 5.1, 6.1, 9.1, 10.1, 11.1, 12.1, 14.1, 15.1, 16.1, where  $i.j$  means  $j$ th replication of the  $i$ th individual), five points under the Student-t distribution (2.1, 3.1, 5.1, 6.1, 16.1) and four points under the Power-Exponential distribution (2.1, 3.1, 6.1, 16.1).

For the response perturbation, Russo *et al.* (2009) found that the measures with the largest influence under normal model are 4.4, 1.7, 12.6, 1.5, 4.7, 7.4, 12.7, 9.7, 8.7 and 10.5. Under Student-t model they are 2.1, 6.7, 9.7, 16.5, 3.5 and 19.7 and finally under Power-Exponential they are 2.1, 6.7, 9.7 and 12.6. As can be seen, these points are not leverage points. Therefore, although the relation between the local influence under response perturbation and the general leverage, the matrices  $\mathbf{D}_\theta$  and  $\boldsymbol{\Delta}$  contain different information about the data.

## 3. Simulation study

In this section we conduct a small simulation study to show some advantages and disadvantages of using the matrix notation. We use the software R Development Core Team (2009) running in a computer Intel

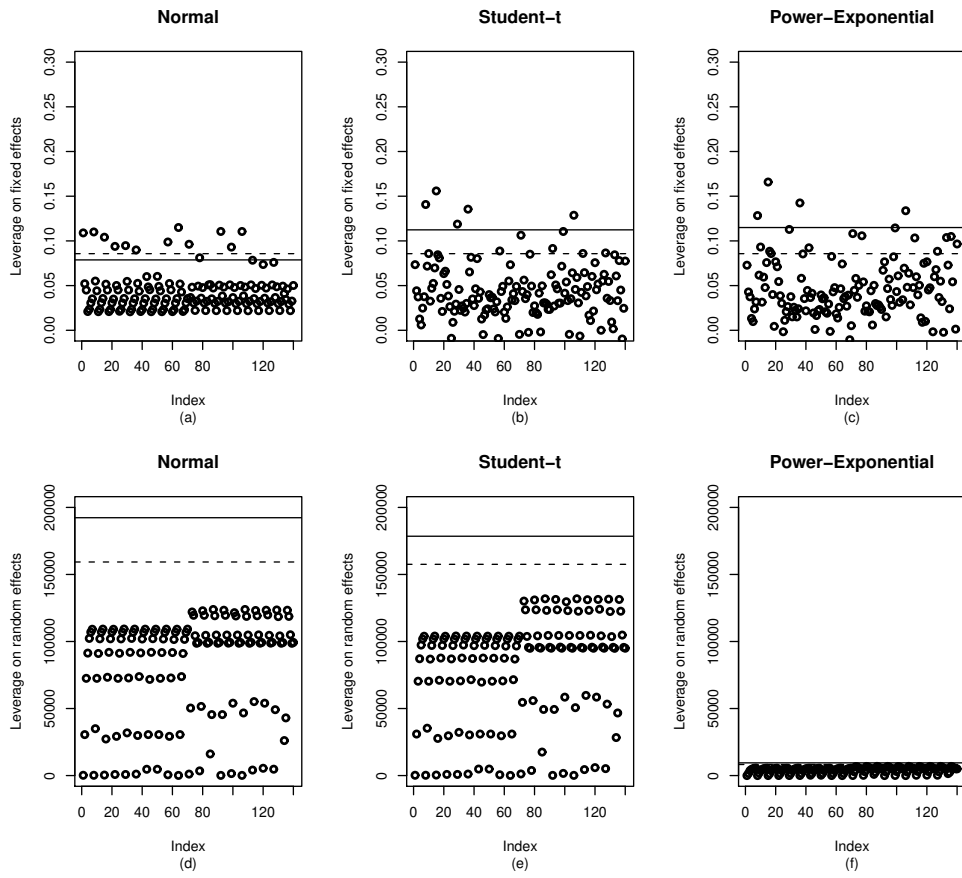


Figure 1: Index plots of leverages on the fixed effects  $\hat{\theta}$  (considering (a) Normal, (b) Student-t and (c) Power-Exponential distributions) and random effects (considering (d) Normal, (e) Student-t and (f) Power-Exponential distributions). The dashed line is  $2tr\{GL/N\}$  and the full line is  $Q_3 + 1.5(Q_3 - Q_1)$ , where  $Q_3$  is the third quartile and  $Q_1$  is first quartile of the diagonal elements of  $GL$ .



Core 2 CPU, 2.14 GHz with 1 Gb of RAM. We use only the formula of the expected Fisher information. We use the following three versions of the Fisher information

$$\mathbf{K}_{\alpha\alpha} \quad \text{and} \quad \mathbf{K}_{\gamma\gamma} = \sum_{i=1}^n \mathbf{V}_i^\top \left[ \frac{c_i}{2} \boldsymbol{\Sigma}_i^{-1} \otimes \boldsymbol{\Sigma}_i^{-1} + (1 - c_i) \text{vec}(\boldsymbol{\Sigma}_i^{-1}) \text{vec}(\boldsymbol{\Sigma}_i^{-1})^\top \right] \mathbf{V}_i, \quad (12)$$

$$\mathbf{K}_{\theta\theta} = \sum_{i=1}^n \mathbf{F}_i^\top \widetilde{\mathbf{H}}_i \mathbf{F}_i. \quad (13)$$

and

$$\mathbf{K}_{\alpha\alpha} \quad \text{and} \quad \mathbf{K}_{\gamma_r\gamma_s} = \sum_{i=1}^n \left\{ \frac{a_{rsi}}{4} (c_i - 1) + c_i \frac{1}{2} \text{tr}[\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(r)} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i(s)}] \right\} \quad \text{for } k, s = 0, \dots, q \quad (14)$$

where all quantities are defined previously. A question that may arise is: when the matrix formulation ((12) and (13)) is faster than the element-by-element formulation (14)? To answer this question, we generate matrices and compute formulas (12), (13) and (14). We consider  $\mathbf{V}_i = \{1/100\}$ ,  $\mathbf{D}_i = \{1/10\}$  and  $\boldsymbol{\Sigma}_i$ , for  $i = 1, \dots, n$ , full non-structured matrices generated from a Wishart distribution for each  $i$  (i.e., these matrices vary with  $i$ ). Notice that the constants  $c_i$ , for  $i = 1, \dots, n$  and the way of generating such matrices are not important, which is important here is the time spent to compute the above formulas. We consider  $n = 20$ ,  $m \in \{7, 14, 21, 28, 35\}$  and  $q \in \{3, 6, 10, 15, 21\}$ . The values of  $q$  were chosen to match with the number of parameters in a non-structured matrix, e.g., for  $q = 3$  we have a  $2 \times 2$  matrix, for  $q = 6$  we have a  $3 \times 3$  matrix and so forth. In the dataset analysed in Russo *et al.* (2009), these values were  $n = 20$ ,  $m = 7$  and  $q = 3$ .

We generate 1000 Monte Carlo samples and compute the mean of the time spent (in seconds) for computing each case. The standard deviations are around  $10^{-5}$  for all cases. Table 1 depicts the average time (in seconds) to compute: (a) formula (12), (b) formula (13) and (c) formula (14). Table 2 presents the ratios: (d) between the figures of (a) and (b) of Table 1; (e) between the figures of (c) and (b) of Table 1 and (f) between the figures of (c) and (a) of Table 1. As expected, when  $q$  increases, the matrix formulations (12) and (13) become faster than the element-by-element one. In particular, if in Russo *et al.* (2009) dataset the matrix  $\mathbf{D}$  had dimensions  $5 \times 5$  ( $q = 21$ ), then the matrix formulation is around 115 (using formula (12)) times faster than the element-by-element one. On the one hand, the time spent to compute the matrix formulation seems to increase very slowly with  $q$ , while for the element-by-element approach this increasing is more pronounced. On the other hand, the time spent for computing the element-by-element formulation seems to increase slowly with  $m$ , while for the matrix approach this increasing is more pronounced. Moreover, as expected, computing  $\mathbf{K}_{\alpha\alpha}$  and  $\mathbf{K}_{\gamma\gamma}$  in (12) is faster than computing  $\mathbf{K}_{\theta\theta}$  directly in (13).

Table 1: Time in seconds: (a) for computing formula (13). (b) for computing formula (14) and (c) for computing formula (12).

	(a)					(b)					(c)				
	$m$					$m$					$m$				
	7	14	21	28	35	7	14	21	28	35	7	14	21	28	35
3	0.0*	0.0*	0.2	0.7	1.8	0.0*	0.2	0.7	2.5	6.5	0.1	0.1	0.1	0.2	0.3
6	0.0*	0.1	0.2	0.7	1.9	0.0*	0.1	0.7	2.4	6.5	0.2	0.3	0.4	0.7	1.1
$q$ 10	0.0*	0.1	0.3	0.8	2.1	0.0*	0.2	0.8	2.4	6.6	0.4	0.8	1.3	2.2	3.4
15	0.0*	0.1	0.3	0.9	2.4	0.0*	0.2	0.8	2.6	6.9	0.9	1.9	3.4	5.9	9.2
21	0.0*	0.1	0.3	1.0	2.6	0.0*	0.2	0.8	2.7	7.2	2.0	4.3	8.2	14.0	21.8

\* lesser than 0.05

#### 4. Final remarks and conclusions

In this short communication, we presented a matrix formulation of the score function, the (expected and observed) Fisher information, the generalized leverage and the  $\boldsymbol{\Delta}$  matrices under case weight, scale and

Table 2: Ratios: (d) the figures of (a) divided by (b); (e) the figures of (c) divided by (b); and (f) the figures of (c) divided by (a).

	(d)					(e)					(f)				
	$m$					$m$					$m$				
	7	14	21	28	35	7	14	21	28	35	7	14	21	28	35
3	0.8	0.3	0.3	0.3	0.3	1.8	0.5	0.2	0.1	0.0*	2.2	1.6	0.6	0.3	0.2
6	0.4	0.3	0.3	0.3	0.3	5.2	1.7	0.6	0.3	0.2	12.5	5.1	1.9	1.0	0.6
$q$ 10	0.4	0.3	0.4	0.3	0.3	13.6	4.7	1.8	0.9	0.5	35.2	13.5	4.6	2.7	1.6
15	0.5	0.4	0.4	0.4	0.3	27.0	11.2	4.4	2.3	1.3	59.0	29.3	12.1	6.5	3.9
21	0.4	0.4	0.4	0.4	0.4	51.7	23.3	9.8	5.1	3.0	115.5	58.8	24.8	13.3	8.3

\* is equal to 0.046

response perturbations for a very general elliptical model which includes the nonlinear mixed-effects elliptical model proposed in Russo *et al.* (2009). The general expressions derived in this paper can be applied in many other models and they are easily interpretable.

Although the matrix formulation may have a high computational cost when the dimension of the observed variable is large, if a proper matrix software is used, the storage requirements of the presented matrix formulation are quite minimal, since in general we are dealing with symmetric and sparse matrices. For instance, by using the packages `Matrix` or/and `SparseM` of the software R Development Core Team (2009) we can have a significant performance improvements in memory utilization for applications involving large sparse matrices (Koenker and Ng, 2009). Regarding the computational speed, on the one hand, we suggest using the element-by-element approach when the dimension of the observed variable ( $m$ ) is large and  $q$  is relatively small. On the other hand, when the dimension of the observed variable is small and  $q$  is relatively large we suggest using the matrix formulation.

## Acknowledgments

I gratefully acknowledge grants from FAPESP. The author also acknowledge helpful comments and suggestions from two anonymous referees which led to an improved presentation.

## References

- Cook, D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society - Series B*, **48**(2), 133–169.
- Demidenko, E. and Stukel, T.A. (2005). Influence analysis for linear mixed-effects models. *Statistics in Medicine*, **24**, 893–909.
- Koenker, R. and Ng, P. (2009) SparseM: A Sparse Matrix Package for R. Available online at: <http://cran.r-project.org/web/packages/SparseM/vignettes/SparseM.pdf>
- Lange, K.L., Little, R.J.A. and Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**, 881–896.
- Magnus, J. R. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester, 3rd edition.
- Mitchell, A. (1989). The information matrix, skewness tensor and  $\alpha$ -connections for the general multivariate elliptic distribution. *Annals of the Institute of Statistics and Mathematics*, **41**:(2), 289–304.
- Nobre, J.S. Singer, J.M. (2010). Leverage analysis for linear mixed models. *Journal of Applied Statistics*, doi: 10.1080/02664761003759016.
- Nobre, J.S. (2004). Métodos de diagnóstico para modelos lineares mistos. São Paulo: Instituto de Matemática e Estatística, Universidade de São Paulo. M.Sc. dissertation (in Portuguese).
- Osorio, F., Paula, G. A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, **51**, 4354–4368.
- Patriota, A.G. and Lemonte, A.J. (2009). Bias correction in a multivariate normal regression model with general parameterization. *Statistics & Probability Letters*, **79**:(15), 1655–1662
- R Development Core Team (2009). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria

- Russo, C.M., Paula, G.A., Aoki, R. (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis*, **53**(12), 4143–4156.
- Savalli, C., Paula, G. A. and Cysneiros, F. J. A. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling* Inglaterra, **6**(1), 59–76.
- Vonesh, E. F. and Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, **48**, 1–17.
- Wei, B.C. (1998). *Exponential Family Nonlinear Models*. Singapore: Springer.
- Wei, B.-C., Hu, Y.-Q., Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics* **25**, 25–37.