A note on Influence diagnostics in nonlinear mixed-effects elliptical models

Alexandre G. Patriota[a,*]

[a]*Institute of Mathematics and Statistics, University of 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 $\boldsymbol{\Delta}$ 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. The matrix formulation has a considerable advantage, 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 analyzed under the assumption of normality. Russo *et al.* (2009) analyze 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. The first author to compute expressions for the entries of the expected Fisher information in a multivariate elliptical distribution was, perhaps, Mitchell (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 elliptic distribution that can be useful to apply in a multivariate context.

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

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

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

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

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

For the purpose of avoiding numerical integrations, Russo *et al.* (2009) consider the marginal model, that is $\boldsymbol{y}_i \overset{ind}{\sim} \mathrm{El}_{m_i}(\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\alpha}); \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^\top + \sigma^2 \boldsymbol{I}_{m_i}$. The vector of parameters of interest is defined

as $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\gamma}^\top)^\top$, where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_q)^\top$ is the vector of parameters involved in $\boldsymbol{\Sigma}_i$ with, in this case, $\gamma_0 = \sigma^2$. In addition to the authors' suppositions, the functional form of $\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\alpha})$ must be known and twice continuously differentiable with respect to each element of $\boldsymbol{\alpha}$.

In this paper, we consider the following model,

$$\boldsymbol{y}_i \overset{ind}{\sim} \mathrm{El}_{m_i}(\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\alpha}); \boldsymbol{\Sigma}_i(\boldsymbol{w}_i, \boldsymbol{\gamma})), \tag{2}$$

where $\boldsymbol{w}_i$ and $\boldsymbol{x}_i$ may have common components. The functional form of the covariance matrix $\boldsymbol{\Sigma}_i(\boldsymbol{w}_i, \boldsymbol{\gamma})$ 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 $\boldsymbol{w}_i = \boldsymbol{Z}_i$ and $\boldsymbol{\Sigma}_i(\boldsymbol{Z}_i, \boldsymbol{\gamma}) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^\top + \sigma^2 \boldsymbol{I}_{m_i}$. 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 $\boldsymbol{R}_i(\boldsymbol{z}_i, \boldsymbol{\sigma}_2)$ instead of $\sigma^2 \boldsymbol{I}_{m_i}$, where $\boldsymbol{z}_i$ is a vector of extra dispersion covariates. Then, in this context, we have that $\boldsymbol{\Sigma}_i(\boldsymbol{w}_i, \boldsymbol{\gamma}) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^\top + \boldsymbol{R}_i(\boldsymbol{z}_i, \boldsymbol{\sigma})$ and $\boldsymbol{\gamma} = (\boldsymbol{\tau}^\top, \boldsymbol{\sigma}^\top)^\top$, where $\boldsymbol{w}_i = (\boldsymbol{Z}_i^\top, \boldsymbol{z}_i^\top)^\top$, $\boldsymbol{\tau}$ is a $q_1 \times 1$ vector of dispersion parameters involved in $\boldsymbol{D}$ 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, $\boldsymbol{\Sigma}_i(\boldsymbol{w}_i, \boldsymbol{\gamma}) = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^\top + \sigma^2 \boldsymbol{V}(\rho)$, where $\boldsymbol{V}_{rs}(\rho) = \rho^{|r-s|}/(1-\rho)$, then $\boldsymbol{w}_i = \boldsymbol{Z}_i$, $q_2 = 2$ and $\boldsymbol{\gamma} = (\boldsymbol{\tau}, \sigma^2, \rho)^\top$. In general, $\boldsymbol{\Sigma}_i(\boldsymbol{w}_i, \boldsymbol{\gamma})$ may be any structured covariance matrix with properties aforementioned. To keep the same notation, consider $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_q)^\top$, 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

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

where $v_i = -2W_g(u_i)$, $u_i = \boldsymbol{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{r}_i$, $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\alpha})$, $\boldsymbol{J}_i = \partial \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^\top$, $\dot{\boldsymbol{\Sigma}}_{i(j)} = \partial \boldsymbol{\Sigma}_i/\partial \gamma_j$, $W_g(u_i) = d \log g(u_i)/du_i$ and function $g(\cdot)$ is the density generator function with properties defined in Russo $et\ al.$ (2009). Notice that, the score function $\boldsymbol{U}_{\gamma_j}$ has a typographical error. The right form is given by

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

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

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

where

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

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

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

with $c_i = 4 f_{gi}/[m_i(m_i + 2)]$ and the quantities d_{gi}, f_{gi} and a_{rsi} 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), but here matrix $\boldsymbol{\Sigma}_i$ does not have the specific structure as regarded in Russo $et\ al.$ (2009), it is left in the general form. See also that the score function and Fisher information for $\boldsymbol{\gamma}$ are written in an element-by-element form.

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). Additionally, Section 1.3 presents an expression for the generalized leverage in model (2). We do not present an application in this paper, since it can 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 $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ be $n \times n$ matrices, define also $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)$ and $\boldsymbol{C} = (\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_n)$, where $\boldsymbol{a}_i$ and $\boldsymbol{c}_i$ are $n \times 1$ vectors, then

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

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,

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

and $\boldsymbol{V}_i = \left(\text{vec}(\dot{\boldsymbol{\Sigma}}_{i(0)}), \ldots, \text{vec}(\dot{\boldsymbol{\Sigma}}_{i(q)})\right)$, where $\boldsymbol{F}_i$ has rank b (i.e., the functions $\boldsymbol{f}$ 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

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

where

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

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

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

where the quantities with the upper script "(m)" are evaluated at $\widehat{\boldsymbol{\theta}}^{(m)}$, $\widetilde{\boldsymbol{H}}_i = \boldsymbol{H}_i \boldsymbol{O}_i \boldsymbol{H}_i$, $\widetilde{\boldsymbol{u}}_i = \boldsymbol{H}_i^{-1} \boldsymbol{O}_i^{-1} \dot{\boldsymbol{u}}_i + \boldsymbol{F}_i \widehat{\boldsymbol{\theta}}$ and m is the iteration counter. Under normality we have that $c_i = 1$, $\boldsymbol{O}_i = \boldsymbol{H}_i^{-1}$ 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 $-\ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = -\sum_{i=1}^n \ddot{\boldsymbol{L}}_{\boldsymbol{\theta}\boldsymbol{\theta},i}$, with

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

where

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\alpha}\boldsymbol{\alpha},i} = 2\boldsymbol{J}_i^\top \boldsymbol{\Sigma}_i^{-1} \left\{ W_g(u_i)\boldsymbol{\Sigma}_i + 2W_g'(u_i)\boldsymbol{r}_i \boldsymbol{r}_i^\top \right\} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{J}_i - 2W_g(u_i)[\boldsymbol{I}_p \otimes \boldsymbol{r}_i^\top \boldsymbol{\Sigma}_i^{-1}]\boldsymbol{D}_i,$$

3

$$\ddot{\boldsymbol{L}}_{\boldsymbol{\alpha\gamma},i} = (\ddot{\boldsymbol{L}}_{\boldsymbol{\alpha}\gamma_0,i}, \ddot{\boldsymbol{L}}_{\boldsymbol{\alpha}\gamma_1,i}, \ldots, \ddot{\boldsymbol{L}}_{\boldsymbol{\alpha}\gamma_q,i}) \tag{6}$$

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

$$\frac{1}{2}\mathrm{tr}\left\{ \boldsymbol{\Sigma}_i^{-1}\left[\dot{\boldsymbol{\Sigma}}_{i(j)}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_{i(k)} - \dot{\boldsymbol{\Sigma}}_{i(jk)} \right] \right\} + \boldsymbol{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \left\{ W_g'(u_i)\dot{\boldsymbol{\Sigma}}_{i(j)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_{i(k)} - W_g(u_i)\dot{\boldsymbol{\Sigma}}_{i(jk)} \right. \tag{7}$$
$$\left. + W_g(u_i)\dot{\boldsymbol{\Sigma}}_{i(j)}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_{i(k)} + W_g(u_i)\dot{\boldsymbol{\Sigma}}_{i(k)}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_{i(j)} \right\} \boldsymbol{\Sigma}_i^{-1}\boldsymbol{r}_i$$

with

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

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

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

where

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

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

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

a) $E(\boldsymbol{r}_i v_i) = \boldsymbol{0}$,

b) $E(\boldsymbol{r}_i\boldsymbol{r}_i^\top v_i) = \boldsymbol{\Sigma}_i$,

c) $E(\boldsymbol{r}_i\boldsymbol{r}_i^\top v_i^2) = 4d_{gi}/m_i\boldsymbol{\Sigma}_i$

d) $E(\mathrm{vec}(\boldsymbol{r}_i\boldsymbol{r}_i^\top)\boldsymbol{r}_i^\top v_i^2) = \boldsymbol{0}$

e) $E(\mathrm{vec}(\boldsymbol{r}_i\boldsymbol{r}_i^\top)\mathrm{vec}(\boldsymbol{r}_i\boldsymbol{r}_i^\top)^\top v_i^2) = c_i\left(\mathrm{vec}(\boldsymbol{\Sigma}_i)\mathrm{vec}(\boldsymbol{\Sigma}_i)^\top + \boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i + \boldsymbol{P}_i(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i) \right)$, where $\boldsymbol{P}_i$ is a commutation maltrix such that $\mathrm{vec}(\boldsymbol{A}) = \boldsymbol{P}_i\mathrm{vec}(\boldsymbol{A}^\top)$ for any matrix $\boldsymbol{A}$ with appropriated dimensions.

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

1.2. Matrix version for $\boldsymbol{\Delta}$

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 $LD(\boldsymbol{\omega}) = 2\{L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ 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 a $s \times 1$ vector of perturbation restricted in an open set $\boldsymbol{\Omega} \subset \mathbb{R}^s$. It is also defined a vector of no perturbation as $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ 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 $C_{\boldsymbol{\ell}}(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^\top \boldsymbol{\Delta}^\top (\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}})^{-1}\boldsymbol{\Delta}\boldsymbol{\ell}|$ where $\boldsymbol{\Delta} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$, both $\boldsymbol{\Delta}$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}$ are evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Thus, $C_{\boldsymbol{d}_{\max}}$ is twice

the largest eigenvalue of $\boldsymbol{B} = -\boldsymbol{\Delta}^\top \ddot{\boldsymbol{L}}_{\boldsymbol{\theta\theta}}^{-1} \boldsymbol{\Delta}$ and $\boldsymbol{d}_{\max}$ is the corresponding eigenvector. The index plot of $\boldsymbol{d}_{\max}$ 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.

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

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

where the quantities with "$\widehat{}$" are evaluated at $\widehat{\boldsymbol{\theta}}$ and

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

In Russo *et al.* (2009), the $\boldsymbol{\Delta}$ matrix under a case weight perturbation is presented with the same typo of the score function. 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 $\boldsymbol{y}_i$ with $\boldsymbol{y}_i + \boldsymbol{\omega}_i$ in $L_i(\boldsymbol{\theta})$) becomes

$$\boldsymbol{\Delta} = \left(\widehat{\boldsymbol{F}}_1^\top \widehat{\boldsymbol{H}}_1 \widehat{\boldsymbol{G}}_1, \ldots, \widehat{\boldsymbol{F}}_n^\top \widehat{\boldsymbol{H}}_n \widehat{\boldsymbol{G}}_n \right), \tag{10}$$

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

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

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

1.3. Generalized leverage

In this section, we compute the generalized leverage proposed by Wei *et al.* (1998). Let $\boldsymbol{y} = \text{vec}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ and $\boldsymbol{\mu}(\boldsymbol{\alpha}) = \text{vec}(\boldsymbol{f}(\boldsymbol{\alpha}, \boldsymbol{x}_1), \ldots, \boldsymbol{f}(\boldsymbol{\alpha}, \boldsymbol{x}_n))$. The authors have shown that the generalized leverage is obtained by evaluating the $N \times N$ matrix

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

at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, where $\boldsymbol{D}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}(\boldsymbol{\alpha})/\partial\boldsymbol{\theta}^\top$ and $\ddot{\boldsymbol{L}}_{\boldsymbol{\theta Y}} = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{Y}^\top$. The main idea behind the concept of leverage is that of evaluating the influence of $\boldsymbol{Y}_i$ 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

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

Index plots of GL_{ii} may reveal those observations with high influence on their own predicted values.

It is worth emphasizing that other models are special cases of the formulas derived in this paper. One just has to define $\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{w}_i, \boldsymbol{\gamma})$ and find their derivatives. That is, the score vector and the (expected and observed) Fisher information as well as the curvatures and the generalized leverage (when available) of several works are special cases of the proposed matrix formulation (to mention just a few of them, see for instance, Paula *et al.*, 2003; Savalli *et al.*, 2006; Osorio *et al.*, 2007; Paula *et al.*, 2009; Russo *et al.*, 2009).

2. Conclusion

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 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 have advantages for numerical purposes because they require only simple operations on matrices and vectors.

Acknowledgments

References

Cook, D., (1986). Assessment of local influence. *Journal of the Royal Statistical Society - Series B*, **48**(2), 133–169

Magnus, J. R. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester, 3rd edition.

Mitchell, (1989). The information matrix, skewness tensor and *a*-connections for the general multivariate elliptic distribution. *Annals of the Institute of Statistics and Mathematics*, **41**:(2), 289–304.

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

Paula, G. A., Cysneiros, F. J. A. and Galea, M. (2003). Local influence and leverage in elliptical nonlinear regression models, *In: Proceedings of the 18th InternationalWorkshop on Statistical Modelling*, Verbeke, G., Molenberghs, G., Aerts, A. and Fieuws, S. (Eds). Leuven: Katholieke Universiteit Leuven, 361–365

Paula, G. A., Medeiros, M. and Vilca-Labrab, F.E. (2009). Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics & Probability Letters*, **79**:(3), 339-346

Russo, C.M., Paula, G.A., Aoki, R. (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis*, doi:10.1016/j.csda.2009.05.004

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.