Ann. Inst. Statist. Math.
Vol. 52, No. 4, 753–766 (2000)

INFLUENCE DIAGNOSTICS IN THE COMMON CANONICAL VARIATES MODEL

HONG GU AND WING K. FUNG

Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong, China

(Received July 14, 1998; revised June 11, 1999)

Abstract. As a generalization of the canonical correlation analysis to k random vectors, the common canonical variates model was recently proposed based on the assumption that the canonical variates have the same coefficients in all k sets of variables, and is applicable to many cases. In this article, we apply the local influence method in this model to study the impact of minor perturbations of data. The method is non-standard because of the restrictions imposed on the coefficients. Besides investigating the joint local influence of the observations, we also obtain the elliptical norm of the empirical influence function as a special case of local influence diagnostics. Based on the proposed diagnostics, we find that the results of common canonical variates analysis for the female water striders data set is largely affected by omitting just one single observation.

Key words and phrases: Common canonical variates, influence function, local influence, perturbation, restricted likelihood, statistical diagnostic.

1. Introduction

Canonical correlation analysis is a classical multivariate method to measure correlation between two sets of variables. The common canonical variates (CCV) model was recently proposed by Neuenschwander and Flury (1995) as a kind of generalization of the canonical correlation analysis to several sets of variables. Unlike Kettenring's (1971) generalizations which were based on the principle of maximizing some generalized measure of canonical correlation, the CCV model was defined based on the assumption that the canonical variates have the same coefficients in all k sets of variables. Such a way of defining the canonical variates facilitates the interpretations of the canonical correlation analysis results and makes sense in many applications.

Suppose the kp-variate random vector x is partitioned into k subvectors $x^{(1)}, \dots, x^{(k)}$ of dimension p each, and denote the positive definite symmetric covariance matrix of x as $\Psi = (\Psi_{ij})_{i,j=1,\dots,k}$. Similar to the common principal components model for independent or dependent random vectors (Flury, (1984, 1988), Flury and Neuenschwander (1995a)), the common canonical variates model assumes that the canonical variates have the same coefficients in all k sets of variables. This assumption is applicable when all the subvectors $x^{(i)}$ have the same dimension p and measure in some sense the same concepts. Common structures are only imposed on the covariance matrix of the variables. Thus, if x satisfies the common canonical variates model, then there exists a nonsingular matrix Γ of dimension $p \times p$ such that $\Lambda_{ij} = \Gamma^T \Psi_{ij} \Gamma$ is diagonal for all submatrices Ψ_{ij} (See Neuenschwander and Flury (1995), Definition 1.1). It follows that the CCV model can be written as

$$(I_k \otimes \Gamma)^T \Psi (I_k \otimes \Gamma) = \Lambda = (\Lambda_{ij})_{i,j=1,\dots,k}.$$

The column vectors in matrix Γ are normalized to unit length and are called the common canonical variates. Therefore, the parameters of interest in this model are the diagonal matrices $\Lambda_{ij} = \text{diag}(\lambda_{ij,1}, \ldots, \lambda_{ij,p})$, $i, j = 1, \ldots, k$, and the nonsingular matrix $\Gamma = (\gamma_1, \ldots, \gamma_p)$ with columns of unit length. The normal theory maximum likelihood estimators of these parameters can be obtained by solving the likelihood equation system given by (7), (8) and (9) in Neuenschwander and Flury (1995), and the equation system can be worked out by the so-called nonorthogonal FG^+ algorithm given by Flury and Neuenschwander (1995*b*).

As an example, Neuenschwander and Flury (1995) presented the common canonical variates analysis results of the female water striders data which was carried in Flury and Neuenschwander (1995*a*). It could be seen from the value of the log-likelihood ratio statistic and the sensible analysis results that the fit of the CCV model is good for this data set. However, as indicated at the end of their article based on a bootstrap analysis, the stability of the coefficients of the first canonical variate is quite poor due to the fact that the likelihood surface is relatively flat in direction corresponding to the first canonical variate. It is anticipated that individual observations may have large influence on model fitting and estimates of the parameters. Thus it is necessary to derive influence diagnostics as the essential supplements to the common canonical variates analysis. In this paper, we shall develop the local influence approach (Cook (1986)) in the CCV model. Cook's (1986) local influence approach is a popular diagnostic method, but the curvature formula in Cook (1986) cannot be directly used for the CCV model because of the restrictions imposed on the normalization of canonical variates. It is not until recently that Kwan and Fung (1998) and Gu and Fung (1998) derived the generalized local influence formulas which can be applied in the restricted likelihood framework. A review of these results is given in Section 2. They will then be used to derive the local influence in the CCV model in Section 3. In Section 4, the female water striders data set is analyzed by applying our newly developed diagnostics, and some more stable results which are largely different from those given by Neuenschwander and Flury (1995) are obtained by omitting just one observation. Besides, a comparison of the statistical diagnostics based on the deletion influence and those based on the local influence approach is also presented in this example. Finally, several summarizing remarks are provided in Section 5.

2. A review of the generalized local influence approach

The local influence approach proposed by Cook (1986) has been applied to various other models besides the linear model and has become a very common diagnostic method. Some general discussion about the methodology can be found in Lawrance (1991), Schall and Dunne (1992), Billor and Loynes (1993) and Fung and Kwan (1997) among others. However, when there are constraints imposed on the parameters of the model, the log-likelihood, and therefore the likelihood displacement, is defined on a restricted parameter space. Thus the curvature formula in Cook (1986) cannot be directly used in such cases. Recently, this problem was investigated by Kwan and Fung (1998) and Gu and Fung (1998). They derived the generalized local influence formulas under the restricted likelihood framework. A review of these results is briefly given below.

Let $L(\theta)$ denote the log-likelihood corresponding to the postulated model, where θ is a $d \times 1$ vector of unknown parameters. A $q \times 1$ vector ω is used to reflect any well-defined perturbation scheme, which is restricted to some open subset Ω of R^q. Let $L(\theta \mid \omega)$ denote the log-likelihood corresponding to the perturbed model for a given ω in Ω such that $L(\theta \mid \omega_0) = L(\theta)$. When there are r constraints $H_i(\theta) = 0$, $i = 1, \ldots, r$, imposed on the d dimensional parameter θ, the log-likelihood $L(\theta)$ and $L(\theta \mid \omega)$ are defined on a $d-r$ dimensional manifold M in R^d. We further assume that $L(\theta \mid \omega)$ is twice continuously differentiable in (θ^T, ω^T), where θ is restricted on M. Suppose ω varies around ω_0 in some fixed direction l, then ω can be represented by $\omega(a) = \omega_0 + al$, where a represents the perturbation scale and l is a nonzero vector of unit length in R^q.

2.1 *All parameters are of interest*

When all parameters in θ are of interest, the likelihood displacement $LD(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_\omega)]$ is used to calibrate how large the influence of the added perturbation is to the model. Instead of using the geometric normal curvature of the influence graph defined in Cook (1986) to measure the local influence, we will use the Taylor expansion of $LD(\omega)$ for local influence analysis. It is because the geometric normal curvature at the null point ω_0 is used to characterize the local change in the likelihood displacement relative to the small changes of ω in Ω space, which is in fact equivalent to using the Taylor expansion of $LD(\omega)$. Suppose $\hat{\theta}$ and $\hat{\theta}_\omega$ are respectively the maximum likelihood estimators under $L(\theta)$ and $L(\theta \mid \omega)$ over the restricted space M. The first order term of the Taylor expansion of $LD(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_\omega)]$ at ω_0 is zero. Let the Lagrange multiplier function of $L(\theta)$ be denoted as

$$f(\theta, \eta) = L(\theta) + \sum_{i=1}^{r} \eta_i H_i(\theta). \tag{2.1}$$

Then for any $\omega \in \Omega$, $LD(\omega) = 2[f(\hat{\theta}, \hat{\eta}) - f(\hat{\theta}_\omega, \hat{\eta})]$. Thus the Taylor approximation of $LD(\omega)$ around ω_0 is given as

$$-a^2 l^T \left(\frac{\partial \hat{\theta}_\omega^T}{\partial \omega} \bigg|_{\omega_0} \right) \left(\frac{\partial^2 f(\theta, \eta)}{\partial \theta \partial \theta^T} \bigg|_{(\hat{\theta}, \hat{\eta})} \right) \left(\frac{\partial \hat{\theta}_\omega}{\partial \omega^T} \bigg|_{\omega_0} \right) l = -a^2 l^T \Delta_\theta^T f^{\theta\theta} \Delta_\theta l, \tag{2.2}$$

where Δ_θ is the $d \times q$ matrix $\partial^2 f(\theta, \eta \mid \omega)/\partial \theta \partial \omega^T$ evaluated at $\theta = \hat{\theta}$ and $\omega = \omega_0$, and $f^{\theta\theta}$ can be obtained from the partition

$$\ddot{f}^{-1} = \left[\frac{\partial^2 f(\theta, \eta)}{\partial (\theta^T, \eta^T)^T \partial (\theta^T, \eta^T)} \right]^{-1} = \begin{bmatrix} f_{\theta\theta} & f_{\theta\eta} \\ f_{\eta\theta} & f_{\eta\eta} \end{bmatrix}^{-1} = \begin{bmatrix} f^{\theta\theta} & f^{\theta\eta} \\ f^{\eta\theta} & f^{\eta\eta} \end{bmatrix}. \tag{2.3}$$

The results presented in (2.2) and (2.3) are the basic generalized local influence formulas extending Cook's (1986) approach, further details are found in Gu and Fung (1998).

Besides the local influence approach, two other commonly used diagnostics are the empirical influence curves (EICs) and the sample influence curves (SICs) (see for example Cook and Weisberg (1982)). The empirical influence curve of the i-th observation to the parameter estimate $\hat{\theta}$, EIC$(x_i, \hat{\theta})$, is actually a special local influence diagnostic. It is found that under the individual case weight perturbation scheme (See Section 3 for the case-weights perturbation scheme), generally we have EIC $(x_i, \hat{\theta}) = c\partial\hat{\theta}_\omega/\partial\omega_i$, where c is a constant related to the sample size and ω_i is the weight of the i-th observation in the estimate. Thus the diagonal elements of $-c^2 \Delta_\theta^T f^{\theta\theta} \Delta_\theta$ are the elliptical norms of EICs which are scaled by matrix $f_{\theta\theta}$ as noted in (2.2). We shall compare the diagonal elements of $-c^2 \Delta_\theta^T f^{\theta\theta} \Delta_\theta$ with the elliptical norms of the sample influence curves SICs scaled by $f_{\theta\theta}$ in the example given in Section 4. The SICs are evaluated under case deletion, which requires a much larger computational load as compared to the evaluation of local influence and the EICs. From the definitions of EIC and SIC given in Cook and Weisberg (1982), we know the limit of SIC when $n \to \infty$ will be equal to EIC. It will be interesting to know the similarities and differences between these two measures in practice. In addition to revealing influence of individual cases, the local influence diagnostics could also indicate joint local influential effects.

2.2 *Subset of parameters is of interest*

Suppose only θ_1 in the partition $\theta^T = (\theta_1^T, \theta_2^T)$ is of interest, where θ_1 is of dimension d_1 and θ_2 is $(d-d_1)\times 1$. In this situation we consider the likelihood profile displacement $LD_S(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_{1\omega}, \theta_2(\hat{\theta}_{1\omega}))]$, where the function $\theta_2(\theta_1)$ maximizes $L(\theta_1, \theta_2)$ on the $d-r$ dimensional manifold M in R^d for any fixed θ_1, and $\hat{\theta}_{1\omega}$ is determined from the partition $\hat{\theta}_\omega^T = (\hat{\theta}_{1\omega}^T, \hat{\theta}_{2\omega}^T)$. Then the Taylor approximation of $LD_S(\omega)$ can be expressed as

$$-a^2 l^T K_1^T (I_{d_1}, K_2^T) f_{\theta\theta} (I_{d_1}, K_2^T)^T K_1 l \tag{2.4}$$

in which, K_1 is the $d_1 \times q$ matrix $\partial\hat{\theta}_{1\omega}/\partial\omega$ evaluated at $\hat{\theta}_1$ and ω_0, K_2 is the $(d-d_1)\times d_1$ matrix $\partial\theta_2(\theta_1)/\partial\theta_1$ evaluated at $\hat{\theta}_1$, and I_{d_1} is the $d_1 \times d_1$ identity matrix. If we denote

$$\ddot{f} = \begin{bmatrix} f_{11} & f_{12} & f_{1\eta} \\ f_{21} & f_{22} & f_{2\eta} \\ f_{\eta 1} & f_{\eta 2} & 0 \end{bmatrix}, \tag{2.5}$$

where $\ddot{f}$ is evaluated at $(\hat{\theta}, \hat{\eta})$, then when θ_2 is not involved in the restriction conditions which is true for the CCV case, we have

$$LD_S(\omega) \approx -a^2 l^T \Delta_\theta^T \left(f^{\theta\theta} - \begin{bmatrix} 0 & 0 \\ 0 & f_{22}^{-1} \end{bmatrix} \right) \Delta_\theta l. \tag{2.6}$$

Information on more general local influence diagnostic, when only a subset of parameters is of interest, can be found in Gu and Fung (1998).

Similar to that indicated in the last subsection, under the individual case weight perturbation, the i-th diagonal element of the matrix $-c^2 K_1^T (I_{d_1}, K_2^T) f_{\theta\theta} (I_{d_1}, K_2^T)^T K_1$ is actually the elliptical norm of EIC$(x_i, \hat{\theta}_1)$ scaled by matrix $(I_{d_1}, K_2^T) f_{\theta\theta} (I_{d_1}, K_2^T)^T$. Likewise, we also use the partial influence which is the elliptical norm of the sample influence curve SICs of $\hat{\theta}_1$ scaled by matrix $(I_{d_1}, K_2^T) f_{\theta\theta} (I_{d_1}, K_2^T)^T$ to compare with the local influence results. When θ_2 is not involved in the restriction conditions, this likelihood contour matrix has a simple form $f_{11} - f_{12} f_{22}^{-1} f_{21}$.

3. Local influence in CCV model

In this section we will derive the local influence diagnostics under the case-weights perturbation scheme. Other perturbation schemes can be discussed similarly.

Suppose we have a sample $x_1, \ldots, x_N$ $(N = n+1)$, where x_i's are kp dimensional and they are iid $N(\mu, \Psi)$, $i = 1, \ldots, N$. Denote the sample covariance matrix as $S = (S_{ij})_{i,j=1,\ldots,k}$, then $M = nS$ is distributed as Wishart with scale matrix Ψ and n degrees of freedom, $n \geq pk$. Let $\omega = (\omega_1, \ldots, \omega_N)^T$, $\omega_i > 0$, denote the case weights for all observations, $\omega_0 = (1, \ldots, 1)^T$ is the null point, and $\omega = \omega_0 + al$ represents a perturbation along some direction l. Under such simultaneous perturbations, suppose the distribution of x_i is perturbed to $N(\mu, \Psi/\omega_i)$, $(i = 1, \ldots, N)$, the maximum likelihood estimates for μ and Ψ are respectively $\bar{x}_\omega = (\sum_{i=1}^N \omega_i x_i)/(\sum_{i=1}^N \omega_i)$ and $M(\omega)/N = \sum_{i=1}^N \omega_i (x_i - \bar{x}_\omega)(x_i - \bar{x}_\omega)^T/N$, and they are statistically independent. Such perturbation was termed the case-weights perturbation scheme by Cook (1986); see Lawrance (1991) for a different name, the variance perturbation scheme. By a result on Wishart distribution given in Rao ((1973), Section 8b.2(ii)), it could be proved that, under the perturbation scheme, $M(\omega)$ is distributed as $W_{kp}(n,\Psi)$. The CCV model imposes structures on the covariance matrix Ψ (Neuenschwander and Flury (1995)). If we denote $G = I_k \otimes \Gamma$, then under the model, we have $G^T\Psi G = \Lambda = (\Lambda_{ij})_{i,j=1,\dots,k}$ and the log-likelihood function after perturbation is

$$(3.1)\quad L(\Lambda,\Gamma\mid\omega) = \frac{n-kp-1}{2}\log|M(\omega)| - \frac{nkp}{2}\log 2 - \sum_{i=1}^{kp}\log\Gamma\left[\frac{1}{2}(n+1-i)\right] - \frac{kp(kp-1)}{4}\log\pi - \frac{n}{2}[\log|\Lambda| - 2\log|G|] - \frac{1}{2}\operatorname{tr}(\Lambda^{-1}G^T M(\omega)G).$$

If we denote the elements of Λ_{ij} by the $p\times 1$ vector $\lambda_{ij} = (\lambda_{ij,1},\dots,\lambda_{ij,p})^T$, $i,j=1,\dots,k$, then all the parameters in this model can be written in an $p^2 + kp(1+k)/2$ dimensional vector $\theta = ((\operatorname{vec}\Gamma)^T,\lambda^T)^T = ((\operatorname{vec}\Gamma)^T,\lambda_{11}^T,\lambda_{12}^T,\dots,\lambda_{1k}^T,\lambda_{22}^T,\lambda_{23}^T,\dots,\lambda_{2k}^T,\ \dots,\lambda_{kk}^T)^T$. Define the Lagrange multiplier function as

$$(3.2)\qquad f(\Gamma,\Lambda,\eta) = L(\Gamma,\Lambda) + \sum_{i=1}^{p}\eta_i(\gamma_i^T\gamma_i - 1),$$

where $L(\Gamma,\Lambda) \equiv L(\Gamma,\Lambda\mid\omega_0)$ is the unperturbed log-likelihood function and $\eta = (\eta_1,\dots,\eta_p)^T$. The corresponding results when our interest is focused on all parameters in the unknown vector θ or only on the common canonical variates Γ are presented in the next two subsections.

3.1 *All parameters in θ are concerned*

The log-likelihood displacement is given as

$$(3.3)\qquad LD(\omega) = 2[L(\hat\theta) - L(\hat\theta_\omega)]$$

$$(3.4)\qquad = 2[f(\hat\theta,\hat\eta) - f(\hat\theta_\omega,\hat\eta)].$$

From (2.2) and (2.3), we have the generalized local influence at ω_0 as

$$(3.5)\qquad -a^2 l^T\Delta_\theta^T f^{\theta\theta}\Delta_\theta l.$$

To get the Δ_θ and $f^{\theta\theta}$ in (3.5), we need the following results.

LEMMA 3.1. *The first order derivatives of the perturbed log-likelihood function $L(\Gamma,\Lambda\mid\omega)$ with respect to the parameters in (Γ,Λ) have the forms*:

$$(3.6)\quad \text{(a)}\quad L_\Gamma = \frac{\partial L(\Gamma,\Lambda\mid\omega)}{\partial\operatorname{vec}(\Gamma)} = knK_{pp}\operatorname{vec}(\Gamma^{-1}) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}[(\Lambda^{ij}\otimes(M_{ij}(\omega)+M_{ji}(\omega)))\operatorname{vec}(\Gamma)],$$

$$(3.7)\quad \text{(b)}\quad L_{\lambda_{ij,h}} = \frac{\partial L(\Gamma,\Lambda\mid\omega)}{\partial\lambda_{ij,h}} = \begin{cases}\frac{1}{2}[F_{ii,h} - n\lambda^{ii,h}], & i=j\\ F_{ij,h} - n\lambda^{ij,h}, & i\neq j\end{cases}$$

where $F = \Lambda^{-1}G^T M(\omega)G\Lambda^{-1}$, $F_{ij,h}$, $\lambda_{ij,h}$ *and* $\lambda^{ij,h}$ *are respectively the h-th diagonal elements of the ij-th* $p \times p$ *blocks of matrices* F, Λ *and* Λ^{-1}. K_{pp} *is the permutation matrix of order* $p \times p$.

The permutation matrix K_{mn} is defined as $K_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n} E_{ij}(m,n) \otimes E_{ij}^T(m,n)$, where $E_{ij}(m,n)$ is an $m \times n$ matrix with 1 at the ij-th position and 0 otherwise. For properties of the permutation matrix and the rules for differentiation, see Magnus and Neudecker ((1988), p. 46 and Chapter 8) or see Fang and Zhang ((1980), p. 13–20). A brief proof of Lemma 3.1 is given in the Appendix.

From (3.6), we have

$$(3.8)\quad f_\Gamma \,|_{(\hat\Gamma,\hat\Lambda,\hat\eta)} = \left.\frac{\partial f(\Gamma,\Lambda,\eta)}{\partial \mathrm{vec}(\Gamma)}\right|_{(\hat\Gamma,\hat\Lambda,\hat\eta)} = L_\Gamma \,|_{(\hat\Gamma,\hat\Lambda)} + 2\,\mathrm{vec}(\Gamma\,\mathrm{diag}(\eta_1,\ldots,\eta_p))\,|_{(\hat\Gamma,\hat\eta)} = 0.$$

Left multiplying (3.8) by $\hat\gamma_t^T, (t = 1,\ldots,p)$, we may get $\hat\eta = 0$ and those equations presented in Theorem 2.1 in Neuenschwander and Flury (1995) which are used for solving the maximum likelihood estimates $\hat\Gamma$ and $\hat\Lambda$.

Further differentiating the results in Lemma 3.1 with respect to ω, we can get

THEOREM 3.1 *Suppose that the t-th observation* $(t = 1,\ldots,N)$ *is written as* $x_t = (x_{t1}^T,\ldots,x_{tk}^T)^T$, *where* $x_{ti}(i = 1,\ldots,k)$ *is of dimension p, and correspondingly* $\bar x = (\bar x_1^T,\ldots,\bar x_k^T)^T$. *Let*

$$(3.9)\qquad \Delta_\theta = \left.\frac{\partial^2 f(\Gamma,\Lambda,\eta \mid \omega)}{\partial\theta\partial\omega^T}\right|_{\theta=\hat\theta,\omega=\omega_0} = \begin{bmatrix} f_{\Gamma\omega} \\ f_{\lambda\omega} \end{bmatrix}.$$

Then the t-th $(t = 1,\ldots,N)$ *columns of* $f_{\Gamma\omega}$ *and* $f_{\lambda\omega}$ *are respectively*

$$(3.10)\qquad \left.\frac{\partial f_\Gamma}{\partial \omega_t}\right|_{(\hat\Gamma,\hat\Lambda,\omega_0)} = -\sum_{i=1}^{k}\sum_{j=1}^{k} \mathrm{vec}[(x_{ti}-\bar x_i)(x_{tj}-\bar x_j)^T\hat\Gamma\hat\Lambda^{ij}],$$

and

$$(3.11)\qquad \left.\frac{\partial f_{\lambda_{ij,h}}}{\partial \omega_t}\right|_{(\hat\Gamma,\hat\Lambda,\omega_0)} = \begin{cases} \frac{1}{2}\hat F_{ii,h}^t, & i = j \\ \hat F_{ij,h}^t, & i \neq j, \end{cases}$$

where $\hat F_{ij,h}^t$ *is the h-th diagonal element of the ij-th* $p \times p$ *block of matrix* $\hat F^t$ *which is* F^t *evaluated at* $(\hat\Gamma,\hat\Lambda)$, *and* $F^t = \Lambda^{-1}G^T(x_t-\bar x)(x_t-\bar x)^T G\Lambda^{-1}$, $t = 1,\ldots,N$.

The proof of Theorem 3.1 is given in the Appendix.

The term left unknown in (3.5) is $f^{\theta\theta}$, which is the upper-left block of matrix $\ddot f^{-1}$. We partition the matrix $\ddot f$ as

$$\ddot f = \left.\frac{\partial^2 f(\Gamma,\Lambda,\eta)}{\partial((\mathrm{vec}\,\Gamma)^T,\lambda^T,\eta^T)^T\partial((\mathrm{vec}\,\Gamma)^T,\lambda^T,\eta^T)}\right|_{(\hat\Gamma,\hat\Lambda,\hat\eta)} = \begin{bmatrix} f_{\Gamma\Gamma} & f_{\Gamma\lambda} & f_{\Gamma\eta} \\ f_{\lambda\Gamma} & f_{\lambda\lambda} & f_{\lambda\eta} \\ f_{\eta\Gamma} & f_{\eta\lambda} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} f_{\Gamma\Gamma} & f_{\Gamma\lambda_{11}} & \cdots & f_{\Gamma\lambda_{1k}} & f_{\Gamma\lambda_{22}} & \cdots & f_{\Gamma\lambda_{2k}} & \cdots & f_{\Gamma\lambda_{kk}} & f_{\Gamma\eta} \\ f_{\lambda_{11}\Gamma} & f_{\lambda_{11}\lambda_{11}} & \cdots & f_{\lambda_{11}\lambda_{1k}} & f_{\lambda_{11}\lambda_{22}} & \cdots & f_{\lambda_{11}\lambda_{2k}} & \cdots & f_{\lambda_{11}\lambda_{kk}} & f_{\lambda_{11}\eta} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ f_{\lambda_{1k}\Gamma} & f_{\lambda_{1k}\lambda_{11}} & \cdots & f_{\lambda_{1k}\lambda_{1k}} & f_{\lambda_{1k}\lambda_{22}} & \cdots & f_{\lambda_{1k}\lambda_{2k}} & \cdots & f_{\lambda_{1k}\lambda_{kk}} & f_{\lambda_{1k}\eta} \\ f_{\lambda_{22}\Gamma} & f_{\lambda_{22}\lambda_{11}} & \cdots & f_{\lambda_{22}\lambda_{1k}} & f_{\lambda_{22}\lambda_{22}} & \cdots & f_{\lambda_{22}\lambda_{2k}} & \cdots & f_{\lambda_{22}\lambda_{kk}} & f_{\lambda_{22}\eta} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ f_{\lambda_{kk}\Gamma} & f_{\lambda_{kk}\lambda_{11}} & \cdots & f_{\lambda_{kk}\lambda_{1k}} & f_{\lambda_{kk}\lambda_{22}} & \cdots & f_{\lambda_{kk}\lambda_{2k}} & \cdots & f_{\lambda_{kk}\lambda_{kk}} & f_{\lambda_{kk}\eta} \\ f_{\eta\Gamma} & f_{\eta\lambda_{11}} & \cdots & f_{\eta\lambda_{1k}} & f_{\eta\lambda_{22}} & \cdots & f_{\eta\lambda_{2k}} & \cdots & f_{\eta\lambda_{kk}} & 0 \end{bmatrix},$$

where all matrices are evaluated at $(\hat{\Gamma}, \hat{\Lambda}, \hat{\eta})$. To derive the matrix $\ddot{f}$, the next result is fundamental.

LEMMA 3.2. *Suppose Λ is a $kp \times kp$ symmetric matrix with all its $p \times p$ submatrices $\Lambda_{ij} (i, j = 1, \ldots, k)$ being diagonal, then we have*

$$\frac{\partial \Lambda^{-1}}{\partial \lambda_{ij,h}} = \begin{cases} -[\lambda^*_{i,h}(\lambda^*_{i,h})^T] \otimes E_{hh}(p,p), & i = j \\ -[\lambda^*_{i,h}(\lambda^*_{j,h})^T + \lambda^*_{j,h}(\lambda^*_{i,h})^T] \otimes E_{hh}(p,p), & i \neq j, \end{cases} \tag{3.12}$$

where $\lambda^*_{i,h} = (\lambda^{i1,h}, \ldots, \lambda^{ik,h})^T$, $E_{ij}(m,n) = e_i(m)e_j(n)^T$, and $e_i(m)$ is a $m \times 1$ vector with 1 at the i-th position and 0 otherwise.

Using the formula $\partial X^{-1}/\partial x_{ij} = -X^{-1}\partial X/\partial x_{ij} X^{-1}$ (Fang and Zhang (1980), p. 15), the proof of Lemma 3.2 is straightforward.

The corresponding result of each block in $\ddot{f}$ is presented in the following Theorem, and the proofs are given in the Appendix.

THEOREM 3.2. *The $p^2 \times p^2$ matrix $f_{\Gamma\Gamma}$ has formula:*

$$f_{\Gamma\Gamma} = -knK_{pp}[(\hat{\Gamma}^{-1})^T \otimes (\hat{\Gamma}^{-1})] - \frac{1}{2}\sum_{i=1}^k \sum_{j=1}^k [\hat{\Lambda}^{ij} \otimes (M_{ij} + M_{ji})]. \tag{3.13}$$

The $p^2 \times kp(k+1)/2$ matrix $f_{\Gamma\lambda}$ has $k(k+1)/2$ submatrices of dimension $p^2 \times p$, all of which are block diagonal matrices with the diagonal blocks given by the $p \times 1$ vectors

$$f_{\gamma_h \lambda_{tm,h}} = \begin{cases} \dfrac{1}{2}\left[\displaystyle\sum_{i=1}^k \sum_{j=1}^k \hat{\lambda}^{ti,h}\hat{\lambda}^{tj,h}(M_{ij} + M_{ji})\hat{\gamma}_h\right], & t = m \\ \dfrac{1}{2}\left[\displaystyle\sum_{i=1}^k \sum_{j=1}^k (\hat{\lambda}^{ti,h}\hat{\lambda}^{mj,h} + \hat{\lambda}^{mi,h}\hat{\lambda}^{tj,h})(M_{ij} + M_{ji})\hat{\gamma}_h\right], & t \neq m. \end{cases} \tag{3.14}$$

The matrix $f_{\Gamma\eta}$ is a $p^2 \times p$ block diagonal matrix with the diagonal blocks $f_{\gamma_h \eta_h} = 2\hat{\gamma}_h$. For the blocks in $f_{\lambda\lambda}$, we have

$$f_{\lambda_{ij,h}\lambda_{ts,h}} = \begin{cases} \dfrac{n}{2}(\hat{\lambda}^{ti,h})^2 - \hat{\lambda}^{ti,h}\hat{F}_{ti,h}, & i = j, \quad t = s \\ n\hat{\lambda}^{ti,h}\hat{\lambda}^{si,h} - \hat{\lambda}^{ti,h}\hat{F}_{si,h} - \hat{\lambda}^{si,h}\hat{F}_{ti,h}, & i = j, \quad t \neq s \\ n\hat{\lambda}^{ti,h}\hat{\lambda}^{sj,h} + n\hat{\lambda}^{si,h}\hat{\lambda}^{tj,h} - \hat{\lambda}^{ti,h}\hat{F}_{sj,h} - \hat{\lambda}^{si,h}\hat{F}_{tj,h} & \\ -\hat{\lambda}^{tj,h}\hat{F}_{si,h} - \hat{\lambda}^{sj,h}\hat{F}_{ti,h}, & i \neq j, \quad t \neq s \end{cases} \tag{3.15}$$

and finally $f_{\lambda\eta} = 0$. *The matrix* $\hat{F} = \hat{\Lambda}^{-1}\hat{G}^T M\hat{G}\hat{\Lambda}^{-1}$ *where* $\hat{G} = I_k \otimes \hat{\Gamma}$.

With all blocks given above, to conduct local influence analysis we define the influence matrix $-\Delta_\theta^T f^{\theta\theta}\Delta_\theta$ as INF. The eigenvector $l_{\max}$ associated with the largest eigenvalue of matrix INF will indicate the way to perturb the case-weights to obtain the greatest local change in the likelihood displacement. Thus $l_{\max}$ could be used to explore the joint influential effects of the data. Besides $l_{\max}$, if there are other eigenvalues which are not significantly smaller than the largest one, the eigenvectors associated with these large eigenvalues may be informative about the joint influence, and may also be investigated.

As indicated in Section 2, the diagonal elements of influence matrix INF are proportional to the elliptical norms of EICs. Because EICs are good approximations to SICs, we will compare the diagonal element of INF with the elliptical norm of $\mathrm{SIC}(x_i, \hat{\theta})/n$ in the example presented in Section 4.

3.2 *Only parameter* Γ *is concerned*

Since the restrictions imposed on the CCV model involve Γ only, by (2.6), the restricted log-likelihood profile displacement is written as

$$LD_s(\omega) \approx -a^2 l^T \Delta_\theta^T \left(f^{\theta\theta} - \begin{bmatrix} 0 & 0 \\ 0 & f_{\lambda\lambda}^{-1} \end{bmatrix} \right) \Delta_\theta l. \tag{3.16}$$

Therefore, the influence matrix for the subset Γ is

$$\mathrm{INF}_s = -\Delta_\theta^T \left(f^{\theta\theta} - \begin{bmatrix} 0 & 0 \\ 0 & f_{\lambda\lambda}^{-1} \end{bmatrix} \right) \Delta_\theta, \tag{3.17}$$

and similar steps can be taken to conduct the local influence analysis as in the last subsection. The diagonal elements of matrix INF_s are now comparable to the norms of $\mathrm{SIC}(x_i, \hat{\Gamma})/n$, $i = 1, \ldots, N$, and the scale matrix for the elliptical norm in this case is $f_{\Gamma\Gamma} - f_{\Gamma\lambda} f_{\lambda\lambda}^{-1} f_{\lambda\Gamma}$.

4. Example: the female water striders data

Water striders grow in six discrete stages called 'instars'. The length of the femur (F) and the tibia (T) of the hind legs of 88 female water striders were recorded in millimeters (Neuenschwander and Flury (1995)). With indices 1 to 3 denoting the first three instars, $X_i = (F_i, T_i)^T$, $i = 1, 2, 3$, could be assembled into a six-dimensional random vector $X = (X_1^T, X_2^T, X_3^T)^T$. Using log-transformed variables and scaled by 100, the CCV analysis results using the whole data set have been presented in Neuenschwander and Flury (1995). Based on these results and the formulas given above, we can get the local influence diagnostics directly. When both $\hat{\Gamma}$ and $\hat{\Lambda}$ are concerned, the comparison of the diagonal elements of influence matrix which are scaled norms of the EICs and the elliptical norms of the SICs/n is given in Fig. 1. It is noted that the SICs are obtained by deleting each single case and performing the FG^+ algorithm of Flury and Neuenschwander (1995*b*) to recompute the parameter estimates. From Fig. 1, it can be seen that both diagnostics basically provide the same information about the influential effects of individual observations, although EICs tend to be underestimate SICs a little bit at those very influential points. Actually, Fig. 1 reflects a rather common relationship between the deletion and local influence measures when the sample size is reasonably

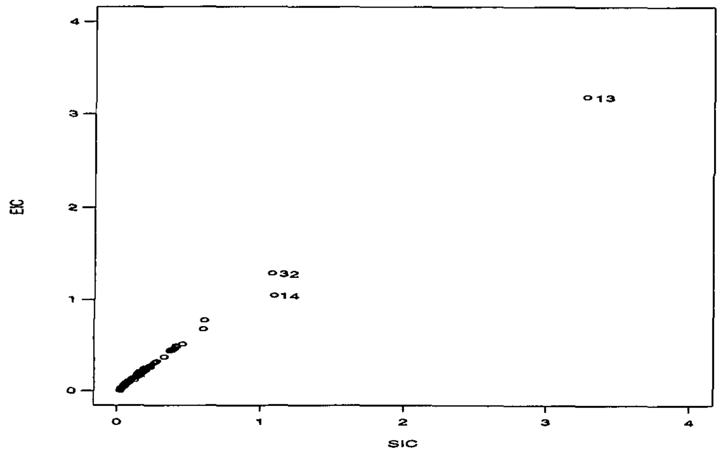

Fig. 1. The comparison of the norms of EICs and SICs when both $\hat{\Gamma}$ and $\hat{\Lambda}$ are of interest.

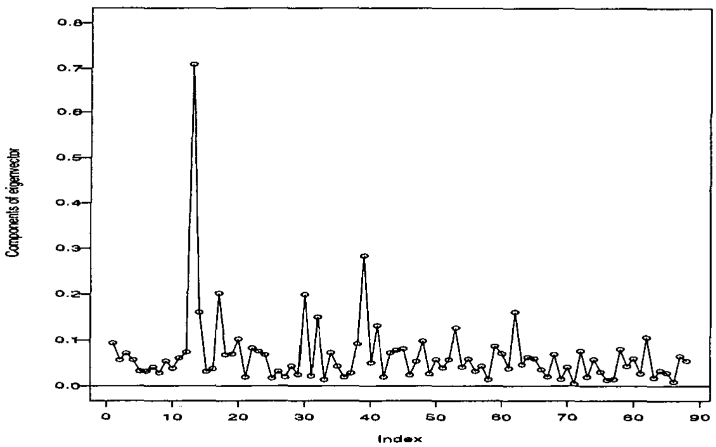

Fig. 2. Index plot of the joint local influence when both $\hat{\Gamma}$ and $\hat{\Lambda}$ are of interest: the components of the first eigenvector of INF.

large. Thus, as an exploratory method to detect which individual observations have undue influence on the parameter estimates, the EICs seems performing quite well.

In addition to the influence of individual observations presented in Fig. 1, the components of the eigenvector associated with the largest eigenvalue of matrix INF plotted in Fig. 2 show the joint effects of these cases. It could be seen from Fig. 2 that case 13 and case 39 have some joint local influence, but it is mainly due to case 13. Figure 1 also shows that case 13 individually has an undue influence to parameter estimation.

As a reference, we also plot in Fig. 3 and Fig. 4 the comparison of local and deletion influence and the components of the first eigenvector of the influence matrix when only $\hat{\Gamma}$ is of interest. In this case the profile log-likelihood displacement and the influence matrix INF_s are used. Figure 4 shows that cases 13 and 39 have large joint local influence to the estimate $\hat{\Gamma}$, and Fig. 3 again reveals that case 13 individually has an undue influence to $\hat{\Gamma}$. When we check the data set carefully, we find that case 13 has the smallest values on all six variables. It is no doubt that case 13 is an influential observation. To see the effect of case 13 on $\hat{\Gamma}$, we give the solutions of $\hat{\Gamma}$ based on the full data set (see also

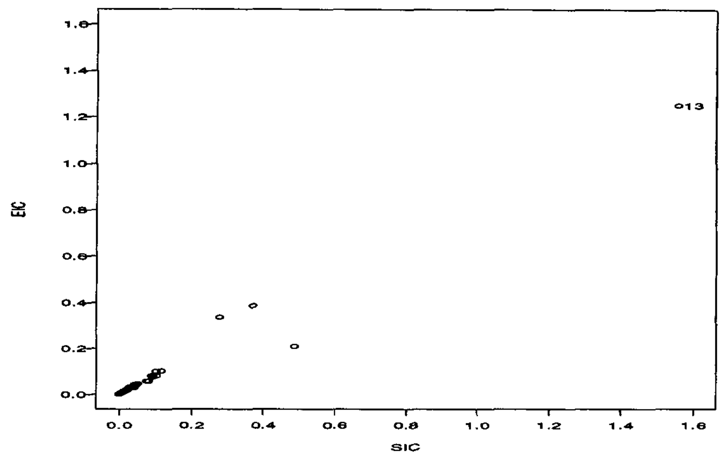

Fig. 3. The comparison of the norms of EICs and SICs when only $\hat{\Gamma}$ is of interest.

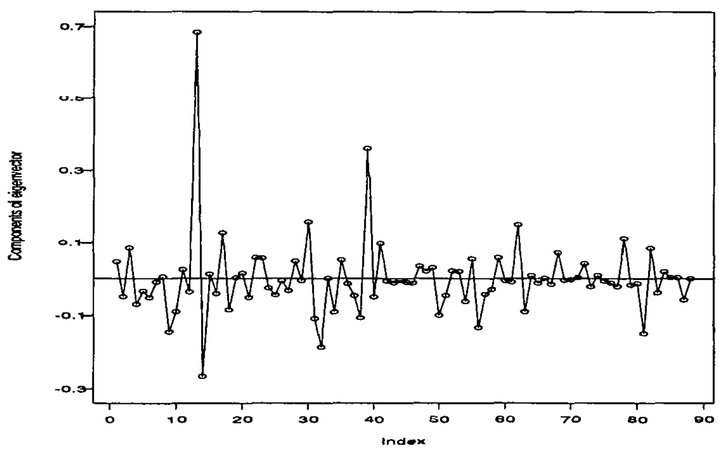

Fig. 4. Index plot of the joint local influence when only $\hat{\Gamma}$ is of interest: the components of the first eigenvector of INF_s.

Neuenschwander and Flury (1995)) and the data set without case 13:

$$\hat{\Gamma} = \begin{pmatrix} 0.731 & 0.763 \\ 0.683 & -0.647 \end{pmatrix}, \quad \hat{\Gamma}_{(13)} = \begin{pmatrix} 0.490 & 0.807 \\ 0.872 & -0.591 \end{pmatrix}.$$

Comparing these two results, it is clear that the estimate for Γ is largely changed by omitting only one case. The matrix $\hat{\Lambda}_{ij}$, $i, j = 1, 2, 3$ could be obtained from the diagonal elements of matrix $\hat{\Gamma}^T S_{ij} \hat{\Gamma}$, $i, j = 1, 2, 3$. We find that if case 13 is omitted, the maximum off-diagonal element of matrix $\hat{\Gamma} S_{ij} \hat{\Gamma}$, $i, j = 1, 2, 3$, is reduced to 0.458 from 0.643 of the full data case, and the convergence speed of the FG^+ algorithm is faster than that of the full data set. The estimate for Λ is also largely affected when case 13 is omitted. For brevity, this result is not shown. The fit of the CCV model could be tested by the usual log-likelihood ratio statistic, which has value 6.65 in the full data case, and value 4.54 if case 13 is deleted. Both values are not significant at any reasonable level comparing to the χ^2 distribution with 7 degrees of freedom, but from the test statistic value, we can still see that the model fits better after case 13 is deleted.

5. Concluding remarks

We have developed the local influence diagnostics in the common canonical variates model. These diagnostics are some essential supplements to the common canonical variates analysis. From these diagnostics, we could not only detect the influential effect of each single case to the model, but also disclose some joint local influence. The derived local influence diagnostics are effective in disclosing the points which have large influence on parameter estimation and on model fitting. As a special part of our local influence diagnostics, the elliptical norm of empirical influence function is obtained. Some comparisons between empirical influence functions and sample influence functions are performed and consistent indications of the influential observations are given by these two measures, which reflects a rather common relationship between the empirical and sample influence functions when the sample size is reasonably large.

Acknowledgements

The authors are most grateful to the editor and the referees for valuable comments which improved the presentation of the paper. The second author is partly supported by the Hong Kong RGC Competitive Earmarked Research Grant HKU 7134/98H.

Appendix

A.1 *Proof of Lemma* 3.1

(a) From (3.1) we have

$$(A.1)\qquad L_\Gamma = n\frac{\partial \log|I_k\otimes\Gamma|}{\partial\,\mathrm{vec}(\Gamma)} - \frac{1}{2}\frac{\partial\,\mathrm{tr}[\Lambda^{-1}(I_k\otimes\Gamma^T)M(\omega)(I_k\otimes\Gamma)]}{\partial\,\mathrm{vec}(\Gamma)}$$
$$= n\frac{\partial k\log|\Gamma|}{\partial\,\mathrm{vec}(\Gamma)} - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial\,\mathrm{tr}[\Gamma^T M_{ij}(\omega)\Gamma\Lambda^{ji}]}{\partial\,\mathrm{vec}(\Gamma)}.$$

By the derivative formula $\partial\,\mathrm{tr}(X^TAXB)/\partial X = AXB + A^TXB^T$ (Fang and Zhang (1980), p. 16), the expression in (A.1) will be

$$knK_{pp}\,\mathrm{vec}(\Gamma^{-1}) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{vec}[M_{ij}(\omega)\Gamma\Lambda^{ji} + M_{ji}(\omega)\Gamma\Lambda^{ij}]$$
$$= knK_{pp}\,\mathrm{vec}(\Gamma^{-1}) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}[(\Lambda^{ij}\otimes(M_{ij}(\omega)+M_{ji}(\omega)))\,\mathrm{vec}(\Gamma)].$$

The result (3.6) of Lemma 3.1 is obtained.

(b) As

$$(A.2)\qquad \frac{\partial\log|\Lambda|}{\partial\lambda_{ij,h}} = \frac{1}{|\Lambda|}\frac{\partial[\mathrm{vec}(\Lambda)]^T}{\partial\lambda_{ij,h}}\frac{\partial|\Lambda|}{\partial\,\mathrm{vec}(\Lambda)} = \frac{\partial[\mathrm{vec}(\Lambda)]^T}{\partial\lambda_{ij,h}}\mathrm{vec}(\Lambda^{-1})$$
$$= \begin{cases}\lambda^{ii,h}, & i=j\\ 2\lambda^{ij,h}, & i\neq j\end{cases}$$

and

$$\text{(A.3)}\quad \frac{\partial \operatorname{tr}[\Lambda^{-1}G^T M(\omega)G]}{\partial \lambda_{ij,h}} = \frac{\partial [\operatorname{vec}(\Lambda^{-1})]^T}{\partial \lambda_{ij,h}} \operatorname{vec}(G^T M(\omega)G)$$
$$= \begin{cases} [\operatorname{vec}(-\Lambda^{-1}E_{ii,h}\Lambda^{-1})]^T \operatorname{vec}[G^T M(\omega)G], & i = j \\ [\operatorname{vec}(-\Lambda^{-1}(E_{ij,h}+E_{ji,h})\Lambda^{-1})]^T \operatorname{vec}[G^T M(\omega)G], & i \neq j \end{cases}$$
$$= \begin{cases} -F_{ii,h}, & i = j \\ -2F_{ij,h}, & i \neq j \end{cases}$$

where $E_{ij,h} = e_{(i-1)p+h}e^T_{(j-1)p+h}$, and $e_{(i-1)p+h}$ is a kp dimensional vector with 1 at the $(i-1)p+h$-th position and 0 otherwise. Substitute (A.2) and (A.3) into

$$\text{(A.4)}\quad \frac{\partial L(\Gamma,\Lambda \mid \omega)}{\partial \lambda_{ij,h}} = -\frac{n}{2}\frac{\partial \log|\Lambda|}{\partial \lambda_{ij,h}} - \frac{1}{2}\frac{\partial \operatorname{tr}[\Lambda^{-1}G^T M(\omega)G]}{\partial \lambda_{ij,h}},$$

with $i = j$ or $i \neq j$ where $i, j = 1, \ldots, k$, (3.7) follows.

A.2 *Proof of Theorem* 3.1

By the result $\partial M_{ij}(\omega)/\partial \omega_t \mid_{\omega_0} = (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)^T$, $(i, j = 1, \ldots, k)$ and Lemma 3.1, it follows that

$$\text{(A.5)}\quad \left.\frac{\partial f_\Gamma}{\partial \omega_t}\right|_{(\hat\Gamma,\hat\Lambda,\omega_0)} = \left.\frac{\partial L_\Gamma}{\partial \omega_t}\right|_{(\hat\Gamma,\hat\Lambda,\omega_0)} = -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \left.\frac{\partial \operatorname{vec}[(M_{ij}(\omega)+M_{ji}(\omega))\hat\Gamma\hat\Lambda^{ij}]}{\partial \omega_t}\right|_{\omega_0}$$
$$= -\sum_{i=1}^{k}\sum_{j=1}^{k} \operatorname{vec}[(x_{ti}-\bar{x}_i)(x_{tj}-\bar{x}_j)^T\hat\Gamma\hat\Lambda^{ij}],$$

and the result in (3.10) is shown. Also by Lemma 3.1, the result in (3.11) is straightforward.

A.3 *Proof of Theorem* 3.2

Let $\operatorname{diag}(\eta_1, \ldots, \eta_p)$ be the $p \times p$ diagonal matrix with diagonal elements $(\eta_1, \ldots, \eta_p)$. Since $f_\Gamma = L_\Gamma + 2\operatorname{vec}[\Gamma \operatorname{diag}(\eta_1, \ldots, \eta_p)]$, we have

$$\text{(A.6)}\quad \frac{\partial f_\Gamma}{\partial (\operatorname{vec}\Gamma)^T}$$
$$= knK_{pp}\frac{\partial \operatorname{vec}(\Gamma^{-1})}{\partial (\operatorname{vec}\Gamma)^T} - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left[(\Lambda^{ij} \otimes (M_{ij}(\omega)+M_{ji}(\omega)))\frac{\partial \operatorname{vec}(\Gamma)}{\partial (\operatorname{vec}\Gamma)^T}\right]$$
$$+ 2\frac{\partial \operatorname{vec}[\Gamma \operatorname{diag}(\eta_1, \ldots, \eta_p)]}{\partial (\operatorname{vec}\Gamma)^T}.$$

By the following results

$$\frac{\partial \operatorname{vec}(\Gamma^{-1})}{\partial (\operatorname{vec}\Gamma)^T} = -(\Gamma^{-1})^T \otimes (\Gamma^{-1}), \quad \frac{\partial \operatorname{vec}(\Gamma)}{\partial (\operatorname{vec}\Gamma)^T} = I_{p^2},$$
$$\frac{\partial \operatorname{vec}[\Gamma \operatorname{diag}(\eta_1, \ldots, \eta_p)]}{\partial (\operatorname{vec}\Gamma)^T} = \operatorname{diag}(\eta_1, \ldots, \eta_p) \otimes I_p,$$

and evaluating (A.6) at $(\hat\Gamma, \hat\Lambda, \hat\eta)$ and ω_0, (3.13) follows.

It is easy to see that $f_{\Gamma\eta} = 2\partial \operatorname{vec}[\Gamma \operatorname{diag}(\eta_1, \ldots, \eta_p)]/\partial\eta^T |_{(\hat{\Gamma},\hat{\Lambda},\hat{\eta})} = b\operatorname{diag}(2\hat{\gamma}_1, \ldots, 2\hat{\gamma}_p)$, where the symbol '$b$ diag' is for the block-diagonal matrix with diagonal blocks as indicated. Since

$$\frac{\partial f_\Gamma}{\partial \lambda_{tm,h}} = \frac{\partial L_\Gamma}{\partial \lambda_{tm,h}} = -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\left(\frac{\partial \Lambda^{ij}}{\partial \lambda_{tm,h}} \otimes (M_{ij}(\omega) + M_{ji}(\omega))\right)\operatorname{vec}(\Gamma)\right],$$

using the result given in Lemma 3.2 and evaluating $\partial f_\Gamma/\partial\lambda_{lm,h}$ at $(\hat{\Gamma}, \hat{\Lambda})$ and ω_0, we obtain

$$\left.\frac{\partial f_\Gamma}{\partial \lambda_{tm,h}}\right|_{(\hat{\Gamma},\hat{\Lambda},\omega_0)} = \begin{cases} -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}[(-\hat{\lambda}^{ti,h}\hat{\lambda}^{tj,h}E_{hh}(p,p)) \otimes (M_{ij}+M_{ji})\operatorname{vec}(\hat{\Gamma})], & t=m \\ -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}[((-\hat{\lambda}^{ti,h}\hat{\lambda}^{mj,h} - \hat{\lambda}^{mi,h}\hat{\lambda}^{tj,h})E_{hh}(p,p)) \\ \qquad\qquad \otimes(M_{ij}+M_{ji})\operatorname{vec}(\hat{\Gamma})], & t\neq m \end{cases}$$

$$= \begin{cases} \frac{1}{2}\operatorname{vec}\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\hat{\lambda}^{ti,h}\hat{\lambda}^{tj,h}(M_{ij}+M_{ji})\hat{\Gamma}E_{hh}(p,p)\right], & t=m \\ \frac{1}{2}\operatorname{vec}\left[\sum_{i=1}^{k}\sum_{j=1}^{k}(\hat{\lambda}^{ti,h}\hat{\lambda}^{mj,h} + \hat{\lambda}^{mi,h}\hat{\lambda}^{tj,h})(M_{ij}+M_{ji})\hat{\Gamma}E_{hh}(p,p)\right], & t\neq m \end{cases}$$

which are equivalent to the results presented in (3.14) (for $E_{hh}(p,p)$, see the definition of $E_{ij}(m,n)$ in Lemma 3.2). Similarly, (3.15) could be derived using Lemma 3.2 and the chain rule, which will not be shown here.

REFERENCES

Billor, N. and Loynes, R. M. (1993). Local influence: a new approach, *Comm. Statist. Theory Methods.*, **22**, 1595–1611.

Cook, R. D. (1986). Assessment of local influence (with discussion), *J. Roy. Statist. Soc. Ser. B*, **48**, 133–169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, London.

Fang, K. T. and Zhang, Y. T. (1980). *Generalized Multivariate Analysis*, Springer, New York.

Flury, B. D. (1984). Common principal components in k groups, *J. Amer. Statist. Assoc.*, **79**, 892–898.

Flury, B. D. (1988). *Common Principal Components and Related Multivariate Models*, Wiley, New York.

Flury, B. D. and Neuenschwander, B. E. (1995*a*). Principal component models for patterned covariance matrices, with applications to canonical correlation analysis of several sets of variables, *Recent Advances in Descriptive Multivariate Analysis*, (Ed. W. J. Krzanowski), 90–112, Clarendon Press, Oxford.

Flury, B. D. and Neuenschwander, B. E. (1995*b*). Simultaneous diagonalization algorithms with applications in multivariate statistics, *Approximation and Computation*, (Ed. R. V. M. Zahar), 179–205, Birkhäuser, Basel.

Fung, W. K. and Kwan, C. W. (1997). A note on local influence based on the normal curvature, *J. Roy. Statist. Soc. Ser. B*, **59**, 839–843.

Gu, H. and Fung, W. K. (1998). Local influence for the restricted likelihood, *Tech. Report*, Department of Statistics and Actuarial Science, The University of Hong Kong.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables, *Biometrika*, **58**, 433–451.

Kwan, C. W. and Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: application to factor analysis, *Psychometrika*, **63**, 35–46.

Lawrance, A. J. (1991). Local and deletion influence, *Direction in Robust Statistics and Diagnostics: Part I* (Eds. W. A. Stahel and S. Weisberg), 141–157, Springer, New York.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.

Neuenschwander, B. E. and Flury, B. D. (1995). Common canonical variates, *Biometrika*, **82**, 553–560.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

Schall, R. and Dunne, T. T. (1992). A note on the relationship between parameter collinearity and local influence, *Biometrika*, **79**, 399–404.