On the asymptotic behavior of general projection–pursuit estimators under the common principal components model*

Graciela Boente[1], Julieta Molina[1] and Mariela Sued[1]

[1] *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*

e–mail: gboente@dm.uba.ar msued@dm.uba.ar

Abstract

The common principal components model for several groups of multivariate observations assumes equal principal axes but possibly different variances along these axes among the groups. The maximum likelihood estimators of the principal axis can be defined as the directions maximizing successively a projection index of the multivariate data defined as the logarithm of the sample variance. Robust estimators can, thus, be defined replacing the classical variability measure by a robust dispersion measure. This projection-pursuit approach was first proposed in Li and Chen (1985) under a principal component analysis and extended to the common principal component setting by Boente, Pires and Rodrigues (2006). The present paper focuses on the consistency and asymptotic distribution of robust projection–pursuit estimators of the common directions under a common principal components model.

Some key words: Asymptotic distribution; Common principal components; Projection–Pursuit; Robust estimation.

Corresponding Author
Graciela Boente
Instituto de Cálculo, FCEyN, UBA
Ciudad Universitaria, Pabellón 2
Buenos Aires, C1428EHA
Argentina
email: gboente@dm.uba.ar
FAX 54-11-45763375

Running Head: Robust common principal component estimators.

*This research was partially supported by Grants X-018 and X-447 from the Universidad de Buenos Aires, PIP 5505, 0592 and 0216 from CONICET and PICT 21407, 00883 and 00821 from ANPCYT, Argentina.

1 Introduction

In many situations, when dealing with several populations, in multivariate analysis, models for common structure dispersion need to be considered to overcome the problem of an excessive number of parameters. Flury (1984) introduced the so–called *Common Principal Components* (CPC) model, in which the common structure assumes that the k covariance matrices have possibly different eigenvalues but identical eigenvectors, i.e.,

$$\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}} , \quad 1 \le i \le k , \tag{1.1}$$

where $\boldsymbol{\Lambda}_i$ are diagonal matrices, $\boldsymbol{\beta}$ is the orthogonal matrix of the common eigenvectors and $\boldsymbol{\Sigma}_i$ is the covariance matrix of the $i-$th population. The maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_i$ are derived in Flury (1984), assuming multivariate normality of the original variables while Flury (1988) considered a unified study of the maximum likelihood estimators under different hierarchical models.

Let $(\mathbf{x}_{ij})_{1\le j\le n_i, 1\le i\le k}$ be independent observations from k independent samples in $\mathbb{R}^p$ with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i$. Let $N = \sum_{i=1}^{k} n_i$, $\tau_{iN} = n_i/N$, where $\tau_{iN} \to \tau_i \in (0,1)$ as $N \to \infty$, and $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})$. For the sake of simplicity and without loss of generality, we will also assume that $\boldsymbol{\mu}_i = \mathbf{0}_p$.

For the CPC model, the common decomposition given in (1.1) implies that for any $\mathbf{a} \in \mathbb{R}^p$, and $1 \le i \le k$, $\mathrm{VAR}\left(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}\right) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}\mathbf{a}$. Therefore, the first (or the last) axis could be obtained through a projection approach by maximizing (or minimizing) $\sum_{i=1}^{k} \tau_i \mathrm{VAR}\left(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}\right)$ over $\mathbf{a} \in \mathbb{R}^p$ with $\|\mathbf{a}\| = 1$. By considering orthogonal directions to $\boldsymbol{\beta}_1$, the second axis is defined and so on. It is well known that, as in the one–population setting, the classical CPC analysis can be affected by the existence of outliers in a sample. The above described projection approach allows to define robust projection–pursuit estimators by considering a robust measure of dispersion (see Boente and Orellana, 2001 and Boente, Pires and Rodrigues, 2002) and provides clear interpretations of the resulting common directions.

On the other hand, it is well known (see Flury (1988)) that the maximum likelihood estimator of $\boldsymbol{\beta}$ for gaussian populations minimizes

$$\prod_{i=1}^{k} \left[\frac{\det\left\{\mathrm{diag}\left(\mathbf{F}_i\right)\right\}}{\det\left(\mathbf{F}_i\right)}\right]^{n_i}$$

where $\mathbf{F}_i = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_i\boldsymbol{\beta}$, and $\mathbf{S}_i$ is the sample covariance matrix of the $i-$th population and so, it can be viewed as the minimizer of $\sum_{j=1}^{p}\sum_{i=1}^{k} n_i \ln\left(\ell_{ij}\right)$, where ℓ_{ij} are the diagonal elements of $\mathbf{F}_i$, i.e., ℓ_{ij} equals the sample variance of the projected vectors $\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}_{i1}, \dots, \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}_{in_i}$. Boente, Pires and Rodrigues (2006) considered a general approach which consists of applying a score function to the scale estimator

Let $f : \mathbb{R}^+ \to \mathbb{R}$ be a general increasing score function and s a univariate scale estimator, Boente, Pires and Rodrigues (2006) propose to estimate the common directions as

$$\begin{cases} \widehat{\boldsymbol{\beta}}_1 & = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_{iN}\, f\left(s^2(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}, \dots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i})\right) \\ \\ \widehat{\boldsymbol{\beta}}_m & = \underset{\mathbf{a}\in\mathcal{B}_m}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_{iN} f\left(s^2(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}, \dots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i})\right) \quad 2 \le m \le p\,; \end{cases} \tag{1.2}$$

where $\mathcal{B}_m = \{\mathbf{a} \in \mathbb{R}^p : \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j = 0 \text{ for } 1 \le j \le m-1\}$. The estimators of the eigenvalues of the i-th population are then computed as $\widehat{\lambda}_{im} = s^2(\widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}\mathbf{x}_{i1}, \dots, \widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}\mathbf{x}_{in_i})$ for $1 \le m \le p$. A different definition arises by minimumizing instead of maximizing, which lead to different solutions (beyond the order) due to the use of a robust scale (see Li and Chen, 1985). However, both proposals will have the same asymptotic behavior.

Partial influence functions of the described propjection–pursuit estimators were derived in Boente, Pires and Rodrigues (2006). The aim of this paper is to obtain under mild conditions the consistency and asymptotic normality of the robust projection–pursuit estimators of the common directions and their size under a CPC model. Asymptotic normality will be derived through Bahadur representations that are applicable to some common choices of robust dispersions. In this sense, our results extend those given by Cui, He and Ng (2003), from one to several populations. Under elliptically symmetric models, our results simplify to provide the same asymptotic variances computed by Boente, Pires and Rodrigues (2006) using partial influence functions.

In Section 2, we describe the general projection–index estimators and the assumptions needed to derive the asymptotic behavior. Our main results are stated in Section 3 where the situation in which all the populations have elliptical distribution except for changes in location and scale is also discussed. Proofs are left to the Appendix. In Boente *et al.* (2006), it was assumed that $n_i = \tau_i N$, where $0 < \tau_i < 1$, are fixed numbers such that $\sum_{i=1}^{k} \tau_i = 1$, i.e., that $\tau_{iN} = n_i/N = \tau_i$. To consider a more general framework, the asymptotic results will be stated by assuming that the sample sizes, n_i, go to infinity in such a way that $\tau_{iN} \to \tau_i \in (0,1)$ and $N^{\frac{1}{2}}(\tau_{iN} - \tau_i) \to 0$. This includes the situation in which n_i is the integer part of $\tau_i N$.

2 Projection–index estimators: Notation and Assumptions

In a one–population setting, robust estimators for the principal directions using alternative measures of variability, were first considered in Li and Chen (1985) who proposed projection pursuit estimators maximizing (or minimizing) a robust scale. Later on, Croux and Ruiz–Gazen (2005) provided the influence functions of the resulting principal components while their asymptotic distribution was studied in Cui, He and Ng (2003).

Under a CPC model, robust projection–pursuit estimators were introduced by Boente and Orellana (2001) who considered as score function f the identity function in (1.2) while in Boente, Pires and Rodrigues (2002) their partial influence function was obtained. Boente, Pires and Rodrigues (2006) proposed the general projection–pursuit estimators defined through (1.2) to estimate the common directions.

From now on $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$ will denote independent vectors from k independent samples in $\mathbb{R}^p$ such that, for $1 \le j \le n_i$ $\mathbf{x}_{ij} \sim F_i$, where F_i is a $p-$dimensional distribution with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i$ satisfying (1.1). As in Boente, Pires and Rodrigues (2006), without loss of generality, we will assume that $\boldsymbol{\mu}_i = 0$. Denote by $F_i[\mathbf{a}]$ the distribution of $\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}$, and by F the product measure $F = F_1 \times F_2 \ldots \times F_k$. Let $\mathcal{F}_1$ be the one dimensional distribution space, $\mathcal{S}_p$ the $p-$dimensional unit sphere and $\mathbf{I}_p$ the identity matrix in $\mathbb{R}^{p\times p}$.

Moreover, let ς be a projection index, i.e., a functional $\varsigma : \mathcal{F}_1 \to \mathbb{R}_{\ge 0}$ and $\sigma(\cdot)$ a univariate scale functional. Denote by $s^2_{i,n_i} : \mathbb{R}^p \to \mathbb{R}$ and $\varsigma_{i,n_i} : \mathbb{R}^p \to \mathbb{R}$ the functions $s^2_{i,n_i}(\mathbf{a}) = \sigma^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$ and $\varsigma_{i,n_i}(\mathbf{a}) = \varsigma(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$, respectively, where $\varsigma(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$ and $\sigma^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$ stand for the functionals ς and σ computed at the empirical distribution of $\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}, \ldots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i}$, respectively. Analogously, $\sigma_i : \mathbb{R}^p \to \mathbb{R}$ and $\varsigma_i : \mathbb{R}^p \to \mathbb{R}$ will stand for $\sigma_i(\mathbf{a}) = \sigma(F_i[\mathbf{a}])$ and $\varsigma_i(\mathbf{a}) = \varsigma(F_i[\mathbf{a}])$, respectively. The estimators defined in Boente, Pires and Rodrigues (2006) correspond to the choice $\varsigma(F) = f(\sigma^2(F))$. We will assume that $\varsigma_i(\mathbf{a}) = \varsigma_i(-\mathbf{a})$ and $\varsigma_{i,n_i}(\mathbf{a}) = \varsigma_{i,n_i}(-\mathbf{a})$, that holds if $\varsigma(F) = f(\sigma^2(F))$. Moreover, let $\rho_N(\mathbf{a})$ and $\rho(\mathbf{a})$ stand for

$$
\begin{aligned}
\rho_N(\mathbf{a}) &= \sum_{i=1}^{k} \tau_{iN}\, \varsigma_{i,n_i}(\mathbf{a}) && (2.1)\\
\rho(\mathbf{a}) &= \sum_{i=1}^{k} \tau_i \varsigma_i(\mathbf{a})\ . && (2.2)
\end{aligned}
$$

Then, a more general framework than (1.2) defines the estimators of the common directions as

$$
\begin{cases}
\widehat{\boldsymbol{\beta}}_1 &= \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_{iN}\varsigma_{i,n_i}(\mathbf{a}) = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}}\, \rho_N(\mathbf{a}) \\
\widehat{\boldsymbol{\beta}}_m &= \underset{\mathbf{a}\in\mathcal{B}_m}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_{iN}\varsigma_{i,n_i}(\mathbf{a}) = \underset{\mathbf{a}\in\mathcal{B}_m}{\operatorname{argmax}}\, \rho_N(\mathbf{a}) \quad 2 \le m \le p.
\end{cases} \tag{2.3}
$$

where $\mathcal{B}_m = \{\mathbf{a} \in \mathbb{R}^p : \ \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j = 0\ ,\ \forall\ 1 \le j \le m-1\}$. The estimators of the eigenvalues of the $i-$th population are then, computed as

$$
\widehat{\lambda}_{im} = \sigma^2(\widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}\mathbf{X}_i) = s^2_{i,n_i}(\widehat{\boldsymbol{\beta}}_m)\ , \quad 1 \le m \le p\ . \tag{2.4}
$$

We will now introduce the statistical functional related to (2.3). Let $\rho(\mathbf{a})$ be defined in (2.2). The projection–index common directions functional $\boldsymbol{\beta}_\varsigma(F) = (\boldsymbol{\beta}_{1,\varsigma}(F), \ldots, \boldsymbol{\beta}_{p,\varsigma}(F))$

is defined as the solution of

$$\begin{cases} \boldsymbol{\beta}_{1,\varsigma}(F) = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}}\, \rho(\mathbf{a}) \\ \\ \boldsymbol{\beta}_{m,\varsigma}(F) = \underset{\mathbf{a}\in\mathcal{C}_m}{\operatorname{argmax}}\, \rho(\mathbf{a}) \quad 2 \le m \le p \end{cases} \tag{2.5}$$

where $\mathcal{C}_m = \{\mathbf{a} \in \mathbb{R}^p : \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_{\ell,\varsigma}(F) = 0\ ,\ \forall\, 1 \le \ell \le m-1\}$. It is clear that both $\widehat{\boldsymbol{\beta}}_m$ and $\boldsymbol{\beta}_{m,\varsigma}(F)$ are defined except for a multiplicative factor -1. As in Cui, He and Ng (2003), we will assume that $\boldsymbol{\beta}_{m,\varsigma}(F)$ is unique, up to direction reversal, for any m, i.e., $\boldsymbol{\beta}_{m,\varsigma}(F)$ and $-\boldsymbol{\beta}_{m,\varsigma}(F)$ are considered equivalent and it does not matter which one we take. In this sense, the convergence $\widehat{\boldsymbol{\beta}}_m \xrightarrow{p} \boldsymbol{\beta}_{m,\varsigma}(F)$ mean convergence in axis, not in the signed vector. In order to identify uniquely the vectors (functional and estimators), one can choose them such that the component with its largest absolute value will be positive, for instance. The eigenvalue functional is defined as

$$\lambda_{im,\varsigma,\sigma}(F) \quad = \quad \sigma^2(F_i[\boldsymbol{\beta}_{m,\varsigma}(F)]) \quad 1 \le m \le p,\ \ 1 \le i \le k\ . \tag{2.6}$$

When $\varsigma(F) = f(\sigma^2(F))$, conditions under which the functional defined through (2.5) will be Fisher–consistent were obtained in Boente, Pires and Rodrigues (2006), while the particular case in which $f(t) = t$ was studied in Boente and Orellana (2001). To simplify the notation, we will avoid the subscript ς and/or σ and so, we will indicate $\boldsymbol{\beta}_m(F) = \boldsymbol{\beta}_{m,\varsigma}(F)$ and $\lambda_{im}(F) = \lambda_{im,\varsigma,\sigma}(F)$.

From now on, the notation $\dot{h}(\mathbf{x}, \mathbf{a})$ will be used for the derivative of the function $h(\mathbf{x}, \mathbf{a})$ with respect to $\mathbf{a}$. Throughout this paper we will consider the following set of assumptions

S0. For some $q \le p$, we have that $\nu_1(F) > \nu_2(F) \ldots > \nu_q(F)$. Moreover, for $1 \le m \le q$, $\boldsymbol{\beta}_m(F)$ are unique except for changes in their sign.

S1. $\varsigma_{i,n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}, \mathbf{a}) + R_{i,n_i}$, where

  a) $\varsigma_i(\mathbf{a})$ is a continuous function of $\mathbf{a}$.
  b) $h_i(\mathbf{x}, \mathbf{a})$ is continuous in both variables.
  c) $\varsigma_{i,n_i}(\mathbf{a})$ is a continuous function of $\mathbf{a}$ a.e.
  d) $\mathrm{E}h_i(\mathbf{x}_{i1}, \mathbf{a}) = 0$ and $\mathrm{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} |h_i(\mathbf{x}_{i1}, \mathbf{a})|\right) < \infty$.
  e) $\sup_{\mathbf{a}\in\mathcal{S}_p} |R_{i,n_i}| \xrightarrow{p} 0$, i.e., $R_{i,n_i} = o_p(1)$ uniformly in $\mathbf{a} \in \mathcal{S}_p$
  f) $\mathrm{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} h_i^2(\mathbf{x}_{i1}, \mathbf{a})\right) < \infty$ and $R_{i,n_i} = o_p\left(n_i^{-\frac{1}{2}}\right)$ uniformly in $\mathbf{a} \in \mathcal{S}_p$.

S2. $\varsigma_i(\mathbf{a})$ is twice continuously differentiable with respect to $\mathbf{a}$. $\dot{\varsigma}_i(\mathbf{a})$ and $\ddot{\varsigma}_i(\mathbf{a})$ will stand for its first and second derivatives, respectively.

S3. The function $\varsigma_{i,n_i}(\mathbf{a})$ is differentiable with respect to $\mathbf{a}$ for any $\mathbf{a} \in \mathcal{S}_p$, almost everywhere. Moreover,

$$\dot{\varsigma}_{i,n_i}(\mathbf{a}) - \dot{\varsigma}_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij}, \mathbf{a}) + \mathbf{o}_p\left(n_i^{-\frac{1}{2}}\right)$$

uniformly in $\mathbf{a} \in \mathcal{S}_p$ with $h_i^\star : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$ such that

a) For any given $\mathbf{x}$, $h_i^\star(\mathbf{x}, \mathbf{a})$ is continuous in $\mathbf{a}$.

b) $\mathrm{E}\, h_i^\star(\mathbf{x}_{i1}, \mathbf{a}) = \mathbf{0}$ for all $\mathbf{a} \in \mathbb{R}^p$ and $\mathrm{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} \|h_i^\star(\mathbf{x}_{i1}, \mathbf{a})\|^2\right) < \infty$.

S4. $s_{i,n_i}(\mathbf{a}) - \sigma_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \mathbf{a}) + R_{i,n_i,\sigma}$, where

a) $\sigma_i(\mathbf{a})$ is a continuous function of $\mathbf{a}$.

b) $\mathrm{E} h_{i,\sigma}(\mathbf{x}_{i1}, \mathbf{a}) = 0$ and $\mathrm{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} |h_{i,\sigma}(\mathbf{x}_{i1}, \mathbf{a})|\right) < \infty$.

c) $h_{i,\sigma}(\mathbf{x}, \mathbf{a})$ is continuous in both variables.

d) $R_{i,n_i,\sigma} = o_p(1)$ uniformly in $\mathbf{a} \in \mathcal{S}_p$

e) $\mathrm{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} h_{i,\sigma}^2(\mathbf{x}_{i1}, \mathbf{a})\right) < \infty$ and $R_{i,n_i,\sigma} = o_p\left(n_i^{-\frac{1}{2}}\right)$ uniformly in $\mathbf{a} \in \mathcal{S}_p$.

S5. The families of functions $\mathcal{H}_i = \{ f(\mathbf{x}) = h_i(\mathbf{x}, \mathbf{a})\ \mathbf{a} \in \mathcal{S}_p\}$,$\mathcal{H}_{i,\sigma} = \{ f(\mathbf{x}) = h_{i,\sigma}(\mathbf{x}, \mathbf{a})\ \mathbf{a} \in \mathcal{S}_p\}$ and $\mathcal{H}_{i,\ell}^\star = \{ f(\mathbf{x}) = h_{i,\ell}^\star(\mathbf{x}, \mathbf{a})\ \mathbf{a} \in \mathcal{S}_p\}$, for $1 \le i \le k$, $1 \le \ell \le p$, with enveloppes $H_i(\mathbf{x}) = \sup_{\mathbf{a}\in\mathcal{S}_p} |h_i(\mathbf{x}, \mathbf{a})|$, $H_{i,\sigma}(\mathbf{x}) = \sup_{\mathbf{a}\in\mathcal{S}_p} |h_{i,\sigma}(\mathbf{x}, \mathbf{a})|$ and $H_{i,\ell}^\star(\mathbf{x}) = \sup_{\mathbf{a}\in\mathcal{S}_p} |h_{i,\ell}^\star(\mathbf{x}, \mathbf{a})|$, respectively, have finite uniform–entropy, where $h_{i,\ell}^\star(\mathbf{x}, \mathbf{a})$ stands for the $\ell-$th component of $h_i^\star(\mathbf{x}, \mathbf{a})$.

Remark 2.1 It is worth noticing that conditions **S1** to **S4** are analoguous to Condtions 1 to 5 in Cui, He and Ng (2003). On the other hand, **S5** is fulfilled if, for instance, $|h_i(\mathbf{x}, \mathbf{a}_1) - h_i(\mathbf{x}, \mathbf{a}_2)| \le G_i(\mathbf{x})\|\mathbf{a}_1 - \mathbf{a}_2\|$ with $\mathrm{E}G_i^2(\mathbf{x}_{i1}) < \infty$, see for instance, van der vaart and Wellner (1996). On the other hand, if $h_i(\mathbf{x}, \mathbf{a}) = \chi_i(\mathbf{a}^{\mathrm{T}}\mathbf{x}/g(\mathbf{a}))$ with $\chi_i : \mathbb{R} \to \mathbb{R}$ a bounded function with bounded variation and $g : \mathbb{R}^p \to \mathbb{R}$, then, **S5** holds. This result follows easily using the permanence properties stated in van der Vaart and Wellner (1996) and that the fact that, given $\epsilon > 0$, for any classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$, if $\mathcal{G} = \{g = g_1 + g_2 : g_i \in \mathcal{G}_i\ ,\ i = 1, 2\}$, then $N\left(\epsilon, \mathcal{G}, L^2(Q)\right) \le N\left(\frac{\epsilon}{2}, \mathcal{G}_1, L^2(Q)\right) \cdot N\left(\frac{\epsilon}{2}, \mathcal{G}_2, L^2(Q)\right)$.

For the sake of simplicity, let us define

$$\nu_m(F) = \max_{\mathbf{a}\in\mathcal{C}_m} \rho(\mathbf{a}) \quad 1 \le m \le p\ , \tag{2.7}$$

$$\widehat{\nu}_m = \max_{\mathbf{a}\in\mathcal{B}_m} \rho_N(\mathbf{a}) \quad 1 \le m \le p\ , \tag{2.8}$$

and let

- $\mathbf{u}_m = -N^{-\frac{1}{2}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m(F))$

- $\mathbf{P}_{m+1} = \mathbf{I}_p - \sum_{j=1}^{m} \boldsymbol{\beta}_j(F)\, \boldsymbol{\beta}_j(F)^{\mathrm{T}}$ the projection matrix over the linear space orthogonal to that spanned by $\boldsymbol{\beta}_1(F), \ldots, \boldsymbol{\beta}_m(F)$,

- $\mathbf{B}_{jm} = \boldsymbol{\beta}_j(F)^{\mathrm{T}}\, \dot{\rho}(\boldsymbol{\beta}_m(F))\, \mathbf{I}_p + \boldsymbol{\beta}_j(F)\, \dot{\rho}(\boldsymbol{\beta}_m(F))^{\mathrm{T}}$,

- $\mathbf{A}_m = \mathbf{P}_{m+1}\, \ddot{\rho}(\boldsymbol{\beta}_m(F)) - \boldsymbol{\beta}_m(F)^{\mathrm{T}}\, \dot{\rho}(\boldsymbol{\beta}_m(F))\mathbf{I}_p - \sum_{j=1}^{m-1} \boldsymbol{\beta}_j(F)^{\mathrm{T}}\, \dot{\rho}(\boldsymbol{\beta}_m(F))\boldsymbol{\beta}_m(F)\, \boldsymbol{\beta}_j(F)^{\mathrm{T}}$.

- With $\mathbf{Z}_0 = 0$ and for $1 \le m \le q$, we define $\mathbf{Z}_m$ recursively by

$$\mathbf{Z}_m = \sum_{j=0}^{m-1} \mathbf{A}_m^{-1} \mathbf{B}_{jm} \mathbf{Z}_j + \mathbf{A}_m^{-1} \mathbf{P}_{m+1} \mathbf{u}_m \,,$$

  provided that $\mathbf{A}_j^{-1}$ exists for $1 \le j \le m$. It is clear that the process $\mathbf{Z}_m$ can by represented by $\mathbf{Z}_m = \sum_{j=0}^{m-1} \mathbf{C}_{jm} \mathbf{u}_j$, for some sequences of matrix $\mathbf{C}_{jm}$ depending on $\mathbf{A}_m$, $\mathbf{B}_{jm}$ and $\mathbf{P}_{m+1}$.

- $\boldsymbol{\xi}_{i,m}(\mathbf{x}) = \sum_{\ell=1}^{m} \mathbf{C}_{\ell m}\, h_i^{\star}(\mathbf{x}, \boldsymbol{\beta}_\ell(F))$, for $\mathbf{x} \in \mathbb{R}^p$.

- $\boldsymbol{\xi}_m(\overrightarrow{\mathbf{x}}) = \sum_{i=1}^{k} \tau_i^{1/2} \boldsymbol{\xi}_{i,m}(\mathbf{x}_i)$, where $\overrightarrow{\mathbf{x}} = (\mathbf{x}_1 \ldots, \mathbf{x}_k)$ and $\mathbf{x}_i \in \mathbb{R}^p$.

3 Main results

3.1 Consistency and Asymptotic Distribution

The following Theorem establishes the consistency of the estimators of the common directions defined through (2.3), under mild conditions. From their consistency, it is easy to derive that of the eigenvalue estimators (2.4) and also that of the estimators of the $i-$scatter matrix defined as $\widehat{\mathbf{V}}_i = \sum_{j=1}^{p} \widehat{\lambda}_{im} \widehat{\boldsymbol{\beta}}_m \widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}$.

Theorem 3.1. *Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$ denote independent vectors from k independent samples in $\mathbb{R}^p$ such that, for $1 \le j \le n_i$, $\mathbf{x}_{ij} \sim F_i$, where F_i is a $p-$dimensional distribution. Moreover, assume that $n_i = \tau_{iN} N$, with $0 < \tau_{iN} < 1$ such that $\sum_{i=1}^{k} \tau_{iN} = 1$ and $\tau_{iN} \to \tau_i \in (0,1)$. Let $\boldsymbol{\beta}_m$, λ_{im} and ν_m be the functionals defined through (2.5), (2.6) and (2.7), respectively. Let $\widehat{\boldsymbol{\beta}}_m$ and $\widehat{\lambda}_{im}$ be the estimators defined in (2.3) and (2.4), respectively. Under **S0**, **S1**a) to e) and **S4**a) to d), we have that, for $1 \le m \le q$, $\widehat{\boldsymbol{\beta}}_m \xrightarrow{p} \boldsymbol{\beta}_m(F)$ and $\widehat{\lambda}_{im} \xrightarrow{p} \lambda_{im}(F)$, for $1 \le i \le k$ as $N \to \infty$.*

The following Theorem gives a Bahadur representation for the estimators $\widehat{\boldsymbol{\beta}}_m$ and $\widehat{\lambda}_{im}$ which allows to derive easily their asymptotic distribution.

Theorem 3.2. *Under the conditions of Theorem 3.1, if, in addition, $N^{\frac{1}{2}}(\tau_{iN}-\tau_i)\to 0$, **S1** to **S5** hold and the matrices $\mathbf{A}_m$, $1\le m\le q$, are non singular, we have that, for $1\le m\le q$,*

$$\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m(F) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\boldsymbol{\xi}_{i,m}(\mathbf{x}_{ij})+\mathbf{o}_p\left(N^{-\frac{1}{2}}\right) \tag{3.1}$$

$$\widehat{\lambda}_{im}-\lambda_{im}(F) = \frac{1}{n_i}\sum_{j=1}^{n_i}h_{i,\sigma}(\mathbf{x}_{ij},\boldsymbol{\beta}_m(F))+o_p\left(n_i^{-\frac{1}{2}}\right) \tag{3.2}$$

Theorem 3.2 entails that, for $1\le i\le k$, the joint distribution of $N^{-\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1(F),\dots,\widehat{\boldsymbol{\beta}}_q-\boldsymbol{\beta}_q(F),\widehat{\lambda}_{i1}-\lambda_{i1}(F),\dots,\widehat{\lambda}_{iq}-\lambda_{iq}(F))$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\text{COV}_F\left(\xi_1(\overrightarrow{\mathbf{x}_1}),\dots,\xi_q(\overrightarrow{\mathbf{x}_1}),\dots,h_{i,\sigma}(\mathbf{x}_{i1},\boldsymbol{\beta}_1(F)),\dots,h_{i,\sigma}(\mathbf{x}_{i1},\boldsymbol{\beta}_q(F))\right)$$

where $\overrightarrow{\mathbf{x}_1}=(\mathbf{x}_{11}\dots,\mathbf{x}_{k1})$.

It is worth noticing that when dealing with only one population, i.e., when $k=1$, Theorem 3.2 provides the Bahadur expansion given in Cui, He and Ng (2003).

3.2 General–Projection pursuit estimates under the CPC model

Let σ be a univariate robust scale functional and $f:\mathbb{R}^+\to\mathbb{R}$ an increasing score function. Considering the functional $\varsigma(\cdot)=f\{\sigma^2(\cdot)\}$ in (2.3), we obtain the estimators defined through (1.2) in Boente, Pires and Rodrigues (2006). As mentioned above, these authors studied conditions for the Fisher–consistency of the common direction functional defined through ς and they also provide an expression for their partial influence functions. From the general results in Pires and Branco (2002), the asymptotic variance of the common direction and eigenvalue estimates, i.e., the variance of the approximating normal distribution, was also computed in Boente, Pires and Rodrigues (2006).

The aim of this section is to show that the expansion obtained Theorem 3.2 allows to obtain under mild conditions the expressions obtained by these authors.

When $\varsigma(\cdot)=f\{\sigma^2(\cdot)\}$, Proposition 1 in Boente, Pires and Rodrigues (2006) provides conditions to ensure the Fisher–consistency of the functionals defined through (2.5). Besides, under regularity conditions on σ, we get that

$$\begin{aligned}
h_{i,\sigma}(\mathbf{x},\mathbf{a}) &= 2\sigma(F_i[\mathbf{a}])\psi_i(\mathbf{x},\mathbf{a})\\
h_i(\mathbf{x},\mathbf{a}) &= f'\left(\sigma^2(F_i[\mathbf{a}])\right)2\sigma(F_i[\mathbf{a}])\psi_i(\mathbf{x},\mathbf{a})\\
h_i^\star(\mathbf{x},\mathbf{a}) &= \dot{h}_i(\mathbf{x},\mathbf{a})=f'\left(\sigma^2(F_i[\mathbf{a}])\right)\left(2\sigma(F_i[\mathbf{a}])\,\psi_i^\star(\mathbf{x},\mathbf{a})+2\dot{\sigma}(F_i[\mathbf{a}])\,\psi_i(\mathbf{x},\mathbf{a})\right)\\
&+ 4\,f''\left(\sigma^2(F_i[\mathbf{a}])\right)\dot{\sigma}(F_i[\mathbf{a}])\sigma^2(F_i[\mathbf{a}])\psi_i(\mathbf{x},\mathbf{a})\quad 1\le i\le k\,,
\end{aligned}$$

where $\psi_i(\mathbf{x},\mathbf{a})=\text{IF}(\mathbf{x},\sigma_{\mathbf{a}};F_i)$, $\psi_i^\star(\mathbf{x},\mathbf{a})=\text{IF}(\mathbf{x},\dot{\sigma}_{\mathbf{a}};F_i)=\dot{\psi}_i(\mathbf{x},\mathbf{a})$ and $\sigma_{\mathbf{a}}:\mathcal{F}_1\to\mathbb{R}^+$ is such that $\sigma_{\mathbf{a}}(F)=\sigma(F[\mathbf{a}])$ and $\dot{\sigma}_{\mathbf{a}}(F)$ is the derivative of $\sigma_{\mathbf{a}}(F)$ respect to $\mathbf{a}$.

Let us consider the following conditions

A1. F_i is an ellipsoidal distribution with location parameter $\boldsymbol{\mu}_i = 0$ and scatter matrix $\boldsymbol{\Sigma}_i = \mathbf{C}_i\mathbf{C}_i^{\mathrm{T}}$ satisfying (1.1). Moreover, when $\mathbf{x}_{ij} \sim F_i$, $\mathbf{C}_i^{-1}\mathbf{x}_{ij} = \mathbf{z}_i$ has the same spherical distribution G for all $1 \le i \le k$.

A2. $\sigma(\cdot)$ is a robust scale functional, equivariant under scale transformations, such that $\sigma(G_0) = 1$, with G_0 the distribution of z_{11}.

A3. The function $(\varepsilon, y) \to \sigma((1-\varepsilon)G_0 + \varepsilon\Delta_y)$ is twice continuously differentiable in $(0, y), y \in \mathbb{R}$ where Δ_y denotes the point mass at y.

A4. f is a twice continuously differentiable function.

A5. For any $1 \le m \le p$, the eigenvalues $\eta_{m\ell} = \sum_{i=1}^k \tau_i f'(\lambda_{im})\lambda_{i\ell}$ of $\widetilde{\boldsymbol{\Sigma}}_m = \sum_{i=1}^k \tau_i f'(\lambda_{im})\boldsymbol{\Sigma}_i$ are such that $\eta_{m\ell} \neq \eta_m = \eta_{mm} = \sum_{i=1}^k \tau_i f'(\lambda_{im})\lambda_{im}$.

Remark 3.1 It is worth noticing that under **A1** to **A4**, the Bahadur expansions required in **S1**, **S3** and **S4** can be obtained under mild conditions on the scale functional, see Cui, He and Ng (2003) for a discussion.

Typically, the influence function of a robust scale functional is bounded. Therefore, using that if **A1** to **A4** hold, $\psi_i(\mathbf{x}, \mathbf{a}) = \sigma_{\mathbf{a}}(F_i)\mathrm{IF}(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_0)$ and $\sigma_{\mathbf{a}}^2(F_i) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a}$ we get easily that $h_{i,\sigma}(\mathbf{x}, \mathbf{a})$ and $h_i(\mathbf{x}, \mathbf{a})$ are bounded. Moreover, if $\mathrm{IF}(y, \sigma; G_0) = \chi(y)$ for some function χ of bounded variation, as is, for instance, the case of an $M-$scale function, then $\mathcal{H}_{i,\sigma}$ and $\mathcal{H}_i$ will have finite uniform–entropy. On the other hand, if $\boldsymbol{\Sigma}_i$ is non-singular, using that

$$\dot{\psi}_i(\mathbf{x}, \mathbf{a}) = \sigma_{\mathbf{a}}^{-1}(F_i)\mathrm{IF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_0\right)\boldsymbol{\Sigma}_i\mathbf{a} + \mathrm{DIF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_0\right)\left(\mathbf{I}_p - \frac{1}{\sigma_{\mathbf{a}}^2(F_i)}\boldsymbol{\Sigma}_i\mathbf{a}\mathbf{a}^{\mathrm{T}}\right)\mathbf{x}$$

where $\mathrm{DIF}(y, \sigma; \mathbf{G}_0)$ denotes the derivative of the influence function $\mathrm{IF}(y, \sigma; \mathbf{G}_0)$ with respect to y. It is easy to see that the first term on the right hand side will be bounded while the second one, can be unbounded for some values of $\mathbf{a}$, for instance, if $\mathbf{a} = \boldsymbol{\beta}_m$. Hence, to ensure that the enveloppe $H_{i,\ell}^{\star}$ has second finite moment it is enough to require that $E\|\mathbf{x}_{i1}\|^2 < \infty$. Therefore, if f' and f'' are functions of bounded variation, to obtain conditions under which $\mathcal{H}_{i,\ell}^{\star}$ will have finite entropy, it is enough to derive conditions ensuring that

$$\mathcal{L}_{i,\ell}^{(1)} = \{L(\mathbf{x}) = f'\left(\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a}\right)\left(\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a}\right)^{\frac{1}{2}}\mathrm{DIF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_0\right)x_\ell\,,\ \mathbf{a} \in \mathcal{S}_p\}$$

$$\mathcal{L}_{i,\ell}^{(2)} = \{L(\mathbf{x}) = f'\left(\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a}\right)\left(\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a}\right)^{\frac{1}{2}}\mathrm{DIF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_0\right)\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}\frac{(\boldsymbol{\Sigma}_i\mathbf{a})_\ell}{\sigma_{\mathbf{a}}(F_i)}\,,\ \mathbf{a} \in \mathcal{S}_p\}$$

have finite entropy. It is worth noticing that $\mathcal{L}_{i,\ell}^{(1)}$ and $\mathcal{L}_{i,\ell}^{(2)}$ will have finite entropy if

$$\mathcal{L}_{i,\ell}^{(1,C)} = \{L(\mathbf{x}) = \lambda\,\mathrm{DIF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_0\right)x_\ell\,,\ \mathbf{a} \in \mathcal{S}_p\,,\ |\lambda| \le C\}$$

$$\mathcal{L}_{i,\ell}^{(2,C)} = \{L(\mathbf{x}) = \lambda\,\mathrm{DIF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_0\right)\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}\,,\ \mathbf{a} \in \mathcal{S}_p\,,\ |\lambda| \le C\}$$

have finite uniform–entropy, which holds for instance if $\mathrm{IF}(y, \sigma; G_0) = \chi(y)$ for some continuously differentiable function χ such that $\chi_1(y) = \chi'(y)$ and $\chi_2(y) = y\chi'(y)$ have bounded variation.

Under **A1** to **A4**, we get that $\boldsymbol{\beta}_m(F) = \boldsymbol{\beta}_m$, $\sigma_i^2(\boldsymbol{\beta}_m) = \lambda_{im}$, $\varsigma_i(\boldsymbol{\beta}_m) = f(\lambda_{im})$, $\dot{\varsigma}_i(\boldsymbol{\beta}_m) = f'(\lambda_{im})\, \lambda_{im}\, \boldsymbol{\beta}_m$ and so,

$$
\begin{aligned}
\psi_i(\mathbf{x}, \boldsymbol{\beta}_m) &= \sqrt{\lambda_{im}}\, \mathrm{IF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma; G_0\right) \\
\dot{\psi}_i(\mathbf{x}, \boldsymbol{\beta}_m) &= \sqrt{\lambda_{im}}\boldsymbol{\beta}_m \mathrm{IF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma;\, G_0\right) + \mathrm{DIF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma;\, G_0\right)(\mathbf{I}_p - \boldsymbol{\beta}_m\boldsymbol{\beta}_m^{\mathrm{T}})\mathbf{x}\,.
\end{aligned}
$$

Moreover, we have that

$$
\begin{aligned}
\rho(\boldsymbol{\beta}_m) &= \sum_{i=1}^{k} \tau_i f(\lambda_{im}) = \nu_m \\
\dot{\rho}(\boldsymbol{\beta}_m) &= 2\eta_m \boldsymbol{\beta}_m \\
\ddot{\rho}(\boldsymbol{\beta}_m) &= 4\sum_{i=1}^{k} \tau_i f''(\lambda_{im})\lambda_{im}^2 \boldsymbol{\beta}_m\boldsymbol{\beta}_m^{\mathrm{T}} + 2\widetilde{\boldsymbol{\Sigma}}_m \\
\dot{h}_i(\mathbf{x}, \boldsymbol{\beta}_m) &= 2\sqrt{\lambda_{im}} f'(\lambda_{im}))\left[\dot{\psi}_i(\mathbf{x}, \boldsymbol{\beta}_m) + \psi_i(\mathbf{x}, \boldsymbol{\beta}_m)\boldsymbol{\beta}_m\right] + 4\, \lambda_{im}^{\frac{3}{2}} f''(\lambda_{im})\, \psi_i(\mathbf{x}, \boldsymbol{\beta}_m)\boldsymbol{\beta}_m\,.
\end{aligned}
$$

Therefore, using that $\boldsymbol{\beta}_j^{\mathrm{T}}\boldsymbol{\beta}_m = 0$ for $1 \le j \le m-1$, we get that

$$
\begin{aligned}
\mathbf{A}_m &= 2\mathbf{P}_{m+1}\widetilde{\boldsymbol{\Sigma}}_m - 2\eta_m\mathbf{I}_p - 2\,\eta_m \sum_{j=1}^{m-1} \boldsymbol{\beta}_j^{\mathrm{T}}\, \boldsymbol{\beta}_m\boldsymbol{\beta}_m\boldsymbol{\beta}_j^{\mathrm{T}} = 2\sum_{i=1}^{k} \tau_i f'(\lambda_{im})(\,\mathbf{P}_{m+1}\,\boldsymbol{\Sigma}_i - \lambda_{im}\mathbf{I}_p) \\
&= 2\sum_{i=1}^{k} \tau_i f'(\lambda_{im})\left(\sum_{j=m+1}^{p} \lambda_{ij}\boldsymbol{\beta}_j\boldsymbol{\beta}_j^{\mathrm{T}} - \lambda_{im}\mathbf{I}_p\right) = 2\sum_{j=m+1}^{p} \eta_{mj}\boldsymbol{\beta}_j\boldsymbol{\beta}_j^{\mathrm{T}} - 2\eta_m\mathbf{I}_p\,,
\end{aligned}
$$

which implies that

$$
\mathbf{A}_m^{-1} = \frac{1}{2}\sum_{j=m+1}^{p} \frac{1}{(\eta_{mj} - \eta_m)}\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}} - \frac{1}{2}\sum_{j=1}^{m} \frac{1}{\eta_m}\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}}. \tag{3.3}
$$

On the other hand, we have that $\mathbf{B}_{jm} = 2\,\eta_m\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_m^{\mathrm{T}}$ for $j \le m-1$, which together with (3.3) leads to $\mathbf{A}_m^{-1}\mathbf{B}_{jm} = -\boldsymbol{\beta}_j\,\boldsymbol{\beta}_m^{\mathrm{T}}$, for $j \le m-1$. Hence, we have that

$$
\mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\mathbf{u}_m = -N^{-\frac{1}{2}} \sum_{i=1}^{k}\sum_{j=1}^{n_i} \sqrt{\lambda_{im}} f'(\lambda_{im})\, \mathrm{DIF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma;\, G_0\right) \sum_{\ell=m+1}^{p} \frac{1}{(\eta_{m\ell} - \eta_m)}(\mathbf{x}_{ij}^{\mathrm{T}}\,\boldsymbol{\beta}_{m,\ell})\,\boldsymbol{\beta}_\ell\,.
$$

From (3.1), we get that

$$
\begin{aligned}
N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) \quad &= \quad -\sum_{\ell=1}^{m-1} \boldsymbol{\beta}_\ell \, \boldsymbol{\beta}_m^{\mathrm{T}} \, N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell) \\
& -N^{-\frac{1}{2}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \sqrt{\lambda_{im}} f'(\lambda_{im}) \, \mathrm{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \, \sigma; \, G_0 \right) \sum_{\ell=m+1}^{p} \frac{1}{(\eta_{m\ell} - \eta_m)} (\mathbf{x}_{ij}^{\mathrm{T}} \, \boldsymbol{\beta}_\ell) \, \boldsymbol{\beta}_\ell + \mathbf{o}_p(1) \; .
\end{aligned}
$$

Therefore, we get that

$$
\begin{aligned}
N^{\frac{1}{2}} \, (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) &= N^{-\frac{1}{2}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \sum_{\ell=1}^{m-1} \sqrt{\lambda_{i\ell}} f'(\lambda_{i\ell}) \, \mathrm{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta}_\ell}{\sqrt{\lambda_{i\ell}}}, \sigma; G_0 \right) \frac{1}{(\eta_{\ell m} - \eta_\ell)} (\mathbf{x}_{ij}^{\mathrm{T}} \, \boldsymbol{\beta}_m) \, \boldsymbol{\beta}_\ell \\
+ \quad N^{-\frac{1}{2}} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \sum_{l=m+1}^{p} & \sqrt{\lambda_{im}} f'(\lambda_{im}) \, \mathrm{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \, \sigma; G_0 \right) \frac{1}{(\eta_m - \eta_{m\ell})} (\mathbf{x}_{ij}^{\mathrm{T}} \, \boldsymbol{\beta}_\ell) \, \boldsymbol{\beta}_\ell + \mathbf{o}_p(1),
\end{aligned}
$$

as suggested by the partial influence functions obtained in Boente, Pires and Rodrigues (2006) and the expansion given in Pires and Branco (2002).

In particular, when $f(t) = t$, we obtain the Bahadur representation of the estimators defined in Boente and Orellana (2001), suggested by the partial influence functions derived in Boente, Pires and Rodrigues (2002).

A Appendix

From now on, whenever it is needed we understand that $\mathcal{B}_1 = \mathcal{C}_1 = \mathcal{S}_p$.

A.1 Proof of Theorem 3.1

The following two Lemmas will be useful to derive the consistency of the estimators. We omit the proof of the first one since it follows straighforwardly.

Lemma A.1. *Let $f_n : \mathcal{S}_p \to \mathbb{R}$ be a sequence of random continuous functions such that $f_n(\mathbf{a}) = f_n(-\mathbf{a})$ and assume that $\sup_{\mathbf{a} \in \mathcal{S}_p} |f_n(\mathbf{a}) - f(\mathbf{a})| = \mathbf{o}_p(1)$ where $f : \mathcal{S}_p \to \mathbb{R}$ is a continuous function such that $f(\mathbf{a}) = f(-\mathbf{a})$. Denote by $\mathbf{w}_n = \mathrm{argmax}_{\mathbf{a} \in \mathcal{S}_p} f_n(\mathbf{a})$ and $\mathbf{w} = \mathrm{argmax}_{\mathbf{a} \in \mathcal{S}_p} f(\mathbf{a})$. Assume that $\mathbf{w}$ is the unique maximum of f in $\mathcal{S}_p$, except for direction reversal, then $\mathbf{w}_n - \mathbf{w} = \mathbf{o}_p(1)$, i.e., $\mathbf{w}_n - \mathbf{w} \xrightarrow{p} 0$, where the convergence $\mathbf{w}_n \xrightarrow{p} \mathbf{w}$ mean convergence in axis, not in the signed vector. In particular, if we choose $\mathbf{w}_n$ and $\mathbf{w}$ such that their component with larger absolute value will be positive, we have that $\mathbf{w}_n \xrightarrow{p} \mathbf{w}$.*

Remark A.1. It is easy to see that if f is a continuous function over a compact $\mathcal{K}$ with a unique maximum, $\mathbf{w}$, and if $(\mathbf{w}_n)_{n \geq 1}$ is a sequence of random elements taking values in $\mathcal{K}$ such that $f(\mathbf{w}_n) - f(\mathbf{w}) = \mathbf{o}_p(1)$, then $\mathbf{w}_n - \mathbf{w} = \mathbf{o}_p(1)$.

Lemma A.2. *Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})$ denote independent vectors from k independent samples in $\mathbb{R}^p$ such that for $1 \le j \le n_i$, $\mathbf{x}_{ij} \sim F_i$, where F_i is a $p-$dimensional distribution.*

*i) Assume that **S1**b), d) and e) holds, then, we have that $\sup_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| \xrightarrow{p} 0$, when $n_i \to \infty$ in such a way that $\tau_{iN} \to \tau_i \in (0,1)$.*

*ii) Assume that **S4**b) to d) holds, then, we have that, for any $1 \le i \le k$, $\sup_{\mathbf{a}\in\mathcal{S}_p} |s_{i,n_i}(\mathbf{a}) - \sigma_i(\mathbf{a})| \xrightarrow{p} 0$, when $n_i \to \infty$.*

Proof. We only prove i) since the proof of ii) is analogous. Note that since

$$\rho_N(\mathbf{a}) - \rho(\mathbf{a}) = \sum_{i=1}^k \tau_{iN} \left(\varsigma_{i,n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})\right) + \sum_{i=1}^k (\tau_i - \tau_{iN})\varsigma_i(\mathbf{a}) ,$$

and $\tau_{iN} \le 1$, we have that

$$\begin{aligned}
\sup_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| &\le \sum_{i=1}^k \tau_{iN} \sup_{\mathbf{a}\in\mathcal{S}_p} |\varsigma_{i,n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})| + \sum_{i=1}^k |\tau_i - \tau_{iN}| \sup_{\mathbf{a}\in\mathcal{S}_p} |\varsigma_i(\mathbf{a})| \\
&\le \sum_{i=1}^k \sup_{\mathbf{a}\in\mathcal{S}_p} |\varsigma_{i,n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})| + \max_{1\le i\le k} |\tau_i - \tau_{iN}| \max_{1\le i\le k} \sup_{\mathbf{a}\in\mathcal{S}_p} |\varsigma_i(\mathbf{a})| ,
\end{aligned}$$

and so using using that $\tau_{iN} \to \tau_i$ and that $\varsigma_i(\mathbf{a})$ is a continuous functions, from **S1**e) it will be enough to show that for any $1 \le i \le k$

$$\sup_{\mathbf{a}\in\mathcal{S}_p} \frac{1}{n_i} \left| \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}, \mathbf{a}) \right| \xrightarrow{p} 0$$

Let $\mathcal{A}_K = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\| \le K\}$. Given $\eta > 0$, let $K \in \mathbb{N}$, be such that

$$\mathrm{E} \sup_{\mathbf{a}\in\mathcal{S}_p} |h_i(\mathbf{x}_{i1}, \mathbf{a})| I_{\mathcal{A}_K^c}(\mathbf{x}_{i1}) \le \eta \tag{A.1}$$

The uniform continuity of $h_i(\mathbf{x}, \mathbf{a})$ over $\mathcal{A}_K \times \mathcal{S}_p$ entails that there exists $\delta > 0$ such that, for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}_K$ and $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{S}_p$ such that $\|\mathbf{x}_1 - \mathbf{x}_2\| < \delta$ and $\|\mathbf{a}_1 - \mathbf{a}_2\| < \delta$, we have that $|h_i(\mathbf{x}_1, \mathbf{a}_1) - h_i(\mathbf{x}_2, \mathbf{a}_2)| < \eta$. Let $(\mathcal{V}_s)_{1\le s\le \ell}$ be a finite collection of balls centered at points $\mathbf{a}_s \in \mathcal{S}_p$ with radius smaller than δ such that $\mathcal{S}_p \subset \bigcup_{s=1}^{\ell} \mathcal{V}_s$, then

$$\max_{1\le s\le \ell} \sup_{\mathbf{a}\in\mathcal{V}_s} \frac{1}{n_i} \left| \sum_{j=1}^{n_i} \left(h_i(\mathbf{x}_{ij}, \mathbf{a}) - h_i(\mathbf{x}_{ij}, \mathbf{a}_s)\right) I_{\mathcal{A}_K}(\mathbf{x}_{ij}) \right| < \eta$$

and so,

$$\sup_{\mathbf{a}\in\mathcal{S}_p} \frac{1}{n_i} \left| \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}, \mathbf{a}) \right| \le \frac{1}{n_i} \sum_{j=1}^{n_i} \sup_{\mathbf{a}\in\mathcal{S}_p} |h_i(\mathbf{x}_{ij}, \mathbf{a})| I_{\mathcal{A}_K^c}(\mathbf{x}_{ij}) + \max_{1\le s\le \ell} \frac{1}{n_i} \left| \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}, \mathbf{a}_s) I_{\mathcal{A}_K}(\mathbf{x}_{ij}) \right| + \eta$$

The law of large numbers and (A.1) entail that

$$A_{n_i}(\eta) = P\left(\frac{1}{n_i}\sum_{j=1}^{n_i}\sup_{\mathbf{a}\in\mathcal{S}_p}|h_i(\mathbf{x}_{ij},\mathbf{a})|\, I_{\mathcal{A}_K^c}(\mathbf{x}_{ij}) > 2\eta\right) \to 0\ .$$

Therefore, the proof will be concluded if we show that

$$B_{n_i}(\eta) = P\left(\max_{1\le s\le \ell}\frac{1}{n_i}\left|\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a}_s) I_{\mathcal{A}_K}(\mathbf{x}_{ij})\right| > 3\eta\right) \to 0\ .$$

Note that

$$\begin{aligned}
B_{n_i}(\eta) &\le \ell\max_{1\le s\le \ell} P\left(\frac{1}{n_i}\left|\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a}_s) I_{\mathcal{A}_K}(\mathbf{x}_{ij})\right| > 3\eta\right)\\
&\le \ell\max_{1\le s\le \ell} P\left(\frac{1}{n_i}\left|\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a}_s)\right| > \eta\right) + \ell\max_{1\le s\le \ell} P\left(\frac{1}{n_i}\left|\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a}_s) I_{\mathcal{A}_K^c}(\mathbf{x}_{ij})\right| > 2\eta\right)\\
&\le \ell\max_{1\le s\le \ell} P\left(\frac{1}{n_i}\left|\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a}_s)\right| > \eta\right) + A_{n_i}(\eta)
\end{aligned}$$

Hence, the proof follows from the fact that $A_{n_i}(\eta)\to 0$ and the law of large numbers using that $\mathrm{E}h_i(\mathbf{x}_{i1},\mathbf{a}_s)=0$, for any $1\le s\le \ell$. $\square$

From now on, to simplify the notation $\boldsymbol{\beta}_j$ and λ_{ij} will stand for $\boldsymbol{\beta}_j(F)$ and $\lambda_{ij}(F)$, respectively

Proof of Theorem 3.1. Let ρ_N, ν_m and $\widehat{\nu}_m$ be defined in (2.1), (2.7) and (2.8) , respectively .

Let us first show that

a) if $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j$, $1\le j\le m-1$, then, $\widehat{\nu}_m \xrightarrow{p} \nu_m$ and

b) if $\widehat{\boldsymbol{\beta}}_m \xrightarrow{p} \boldsymbol{\beta}_m$, then, $\widehat{\lambda}_{im} \xrightarrow{p} \lambda_{im}$, which shows the consistency of the eigenvalue estimators.

a) Lemma A.2 entails that

$$\sup_{\mathbf{a}\in\mathcal{S}_p}|\rho_N(\mathbf{a})-\rho(\mathbf{a})| \xrightarrow{p} 0\ , \tag{A.2}$$

and so

$$\begin{aligned}
|\widehat{\nu}_m-\nu_m| &= |\rho_N(\widehat{\boldsymbol{\beta}}_m)-\rho(\boldsymbol{\beta}_m)| = |\max_{\mathbf{a}\in\mathcal{B}_m}\rho_N(\mathbf{a}) - \max_{\mathbf{a}\in\mathcal{C}_m}\rho(\mathbf{a})|\\
&\le \max_{\mathbf{a}\in\mathcal{S}_p}|\rho_N(\mathbf{a})-\rho(\mathbf{a})| + |\max_{\mathbf{a}\in\mathcal{C}_m}\rho(\mathbf{a}) - \max_{\mathbf{a}\in\mathcal{B}_m}\rho(\mathbf{a})|
\end{aligned}$$

and thus, the result follows from (A.2), the continuity of ρ and the fact that $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j$, $1 \le j \le m-1$.

b) Using Lemma A.2 ii), we get that $\sup_{\mathbf{a}\in\mathcal{S}_p} |s_{i,n_i}(\mathbf{a}) - \sigma_i(\mathbf{a})| \xrightarrow{p} 0$. Hence, since $\widehat{\lambda}_{im} = s^2_{i,n_i}(\widehat{\boldsymbol{\beta}}_m)$, $\lambda_{im} = \sigma_i^2(\boldsymbol{\beta}_m)$ and

$$\begin{aligned} |\widehat{\lambda}_{im} - \lambda_{im}| &= |s^2_{i,n_i}(\widehat{\boldsymbol{\beta}}_m) - \sigma_i^2(\widehat{\boldsymbol{\beta}}_m) + \sigma_i^2(\widehat{\boldsymbol{\beta}}_m) - \sigma_i^2(\boldsymbol{\beta}_m)| \\ &\le \max_{\mathbf{a}\in\mathcal{S}_p} |s^2_{i,n_i}(\mathbf{a}) - \sigma_i^2(\mathbf{a})| + |\sigma_i^2(\widehat{\boldsymbol{\beta}}_m) - \sigma_i^2(\boldsymbol{\beta}_m)| \end{aligned}$$

the continuity of $\sigma_i(\mathbf{a})$ and the fact that $\widehat{\boldsymbol{\beta}}_m \xrightarrow{p} \boldsymbol{\beta}_m$ entail that $\widehat{\lambda}_{im} \xrightarrow{p} \lambda_{im}$.

Therefore, it remains to show that $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j$. From (A.2), using that $\widehat{\boldsymbol{\beta}}_1$ and $\boldsymbol{\beta}_1$ maximize ρ_N and ρ, respectively, the continuity of ρ_N and ρ and Lemma A.1, we obtain easily that $\widehat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\beta}$.

Let $2 \le m \le p$ and assume that $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j$, $1 \le j \le m-1$, we will then show that $\widehat{\boldsymbol{\beta}}_m \xrightarrow{p} \boldsymbol{\beta}_m$. Denote by $\mathbf{a}_m, \dots, \mathbf{a}_p$ an orthonormal basis of the linear space orthogonal to that spanned by $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{m-1}$, then, $\{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{m-1}, \mathbf{a}_m, \dots, \mathbf{a}_p\}$ is an ortonormal basis of $\mathbb{R}^p$. Hence, we have that

$$\widehat{\boldsymbol{\beta}}_m = \sum_{j=1}^{m-1} \langle \widehat{\boldsymbol{\beta}}_m, \boldsymbol{\beta}_j \rangle \boldsymbol{\beta}_j + \sum_{i=m}^{p} \langle \widehat{\boldsymbol{\beta}}_m, \mathbf{a}_i \rangle \mathbf{a}_i = \sum_{j=1}^{m-1} \widehat{c}_j \boldsymbol{\beta}_j + \mathbf{b}_m \ ,$$

where $\langle \boldsymbol{\beta}_j, \mathbf{b}_m \rangle = 0$, for $1 \le j \le m-1$. On the other hand, since $\|\widehat{\boldsymbol{\beta}}_m\| = 1$, $\langle \widehat{\boldsymbol{\beta}}_j, \widehat{\boldsymbol{\beta}}_m \rangle = 0$ and $\widehat{\boldsymbol{\beta}}_j \xrightarrow{p} \boldsymbol{\beta}_j$ for $1 \le j \le m-1$, we have that

$$\widehat{c}_j = \langle \widehat{\boldsymbol{\beta}}_m, \boldsymbol{\beta}_j \rangle = \langle \widehat{\boldsymbol{\beta}}_m, \boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j \rangle \xrightarrow{p} 0 \ ,$$

and so, $\|\mathbf{b}_m\| \xrightarrow{p} 1$. Let $\mathbf{b}_m = \widehat{c}_m \, \widehat{\mathbf{b}}_m$, where $\widehat{c}_m = \|\mathbf{b}_m\| \xrightarrow{p} 1$ and $\|\widehat{\mathbf{b}}_m\| = 1$, then,

$$\widehat{\boldsymbol{\beta}}_m - \widehat{\mathbf{b}}_m = \widehat{\boldsymbol{\beta}}_m - \widehat{\mathbf{b}}_m + \widehat{\mathbf{b}}_m \, \widehat{c}_m - \widehat{\mathbf{b}}_m \, \widehat{c}_m = \widehat{\boldsymbol{\beta}}_m - \mathbf{b}_m + \widehat{\mathbf{b}}_m \, (\widehat{c}_m - 1) = \sum_{j=1}^{m-1} \widehat{c}_j \, \boldsymbol{\beta}_j + \widehat{\mathbf{b}}_m \, (\widehat{c}_m - 1) \xrightarrow{p} 0 \ . \tag{A.3}$$

Therefore, to prove the consistency of $\widehat{\boldsymbol{\beta}}_m$ it is enough to show that $\widehat{\mathbf{b}}_m - \boldsymbol{\beta}_m \xrightarrow{p} 0$. Using that $\widehat{\mathbf{b}}_m$, $\boldsymbol{\beta}_m \in \mathcal{C}_m$, $\boldsymbol{\beta}_m$ is the unique maximum of ρ over $\mathcal{C}_m$ and Remark A.1, it is enough to show that $\rho(\widehat{\mathbf{b}}_m) - \rho(\boldsymbol{\beta}_m) \xrightarrow{p} 0$. Note that

$$\begin{aligned} |\rho(\widehat{\mathbf{b}}_m) - \rho(\boldsymbol{\beta}_m)| &\le |\rho(\widehat{\mathbf{b}}_m) - \rho(\widehat{\boldsymbol{\beta}}_m)| + |\rho(\widehat{\boldsymbol{\beta}}_m) - \rho_N(\widehat{\boldsymbol{\beta}}_m)| + |\rho_N(\widehat{\boldsymbol{\beta}}_m) - \rho(\boldsymbol{\beta}_m)| \\ &\le |\rho(\widehat{\mathbf{b}}_m) - \rho(\widehat{\boldsymbol{\beta}}_m)| + \sup_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| + |\widehat{\nu}_m - \nu_m| \end{aligned}$$

Therefore, $\rho(\widehat{\mathbf{b}}_m) - \rho(\boldsymbol{\beta}_m) \xrightarrow{p} 0$ since ρ is a continuous function, (A.3) and (A.2) hold and and $\widehat{\nu}_m - \nu_m = \mathbf{o}_p(1)$. $\square$

A.2 Proof of Theorem 3.2

To aid the proof of the Bahadur representations in (3.1), we will state several lemmas conducting to the desired conclusion.

Lemma A.3. *Under the conditions of Theorem 3.2, we have that*

i) $\rho_N(\widehat{\boldsymbol\beta}_m) - \rho_N(\boldsymbol\beta_m) - \rho(\widehat{\boldsymbol\beta}_m) + \rho(\boldsymbol\beta_m) = \mathbf{o}_p(N^{-\frac{1}{2}})$

ii) $\dot\rho(\widehat{\boldsymbol\beta}_m) + N^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij},\widehat{\boldsymbol\beta}_m) - \left\{\dot\rho(\boldsymbol\beta_m) + N^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\right\} = \ddot\rho(\boldsymbol\beta_m)(\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_m) + \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_m\|) + \mathbf{o}_P(N^{-\frac{1}{2}})$

Proof. We will only show ii) of the Lemma, the proof of i) being analogous using the fact that $N^{1/2}(\tau_{iN}-\tau_i)\to 0$, $0\le\tau_{iN}\le 1$, the continuity of ς_i, **S1f)** and the equality

$$\rho_N(\mathbf{a}) - \rho(\mathbf{a}) = \sum_{i=1}^k \tau_{iN}\left(\varsigma_{i,n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})\right) + \sum_{i=1}^k(\tau_i - \tau_{iN})\varsigma_i(\mathbf{a})\ ,$$

Given $1\le\ell\le p$, under **S5** and since $E\{h_{i,\ell}^\star(\mathbf{x}_{i1},\mathbf{a})\} = 0$, we have that $\mathcal{H}_{i,\ell}^\star$ is Donsker. Therefore, denoting $d^2(\mathbf{a},\mathbf{b}) = E\left(h_{i,\ell}^\star(\mathbf{x}_{i1},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{i1},\mathbf{b})\right)^2$, we have that (see van der vaart and Wellner, 1996, page 115), for any $\alpha_N\to 0$

$$\sup_{\substack{d(\mathbf{a},\boldsymbol\beta_m)\le\alpha_N\\ \mathbf{a}\in\mathcal{S}_p}} E\left|n_i^{-\frac{1}{2}}\sum_{j=1}^{n_i}\{h_{i,\ell}^\star(\mathbf{x}_{ij},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\}\right| \to 0\ ,\ \text{for } 1\le i\le k\ .$$

Therefore using that $0\le\tau_{iN}\le 1$ and

$$\begin{aligned}
E|N^{-\frac{1}{2}}\sum_{i=1}^k\sum_{j=1}^{n_i}\{h_{i,\ell}^\star(\mathbf{x}_{ij},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\}| &= E|N^{-\frac{1}{2}}\sum_{i=1}^k n_i^{\frac{1}{2}} n_i^{-\frac{1}{2}}\sum_{j=1}^{n_i}\{h_{i,\ell}^\star(\mathbf{x}_{ij},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\}| \\
&\le \sum_{i=1}^k \tau_{iN}^{\frac{1}{2}} E|n_i^{-\frac{1}{2}}\sum_{j=1}^{n_i}\{h_{i,\ell}^\star(\mathbf{x}_{ij},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\}| \\
&\le \max_{1\le i\le k} E|n_i^{-\frac{1}{2}}\sum_{j=1}^{n_i}\{h_{i,\ell}^\star(\mathbf{x}_{ij},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\}|
\end{aligned}$$

we get that

$$\sup_{\substack{d(\mathbf{a},\boldsymbol\beta_m)\le\alpha_N\\ \mathbf{a}\in\mathcal{S}_p}} E\left|N^{-\frac{1}{2}}\sum_{i=1}^k\sum_{j=1}^{n_i}\{h_{i,\ell}^\star(\mathbf{x}_{ij},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\}\right| \to 0\ .$$

The continuity on $\mathbf{a}$ of $h_{i,\ell}^\star(\mathbf{x},\mathbf{a})$ implies that $\lim_{\|\mathbf{a}-\boldsymbol\beta_m\|\to 0} d(\mathbf{a},\boldsymbol\beta_m) = 0$, and so, if $\alpha_N\to 0$

$$\sup_{\substack{\|\mathbf{a}-\boldsymbol\beta_m\|\le\alpha_N\\ \mathbf{a}\in\mathcal{S}_p}} \left|\frac{1}{N}\sum_{i=1}^k\sum_{j=1}^{n_i}\{h_{i,\ell}^\star(\mathbf{x}_{ij},\mathbf{a}) - h_{i,\ell}^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\}\right| = \mathbf{o}_p(N^{-\frac{1}{2}})\ ,$$

which together with the consistency of $\widehat{\boldsymbol{\beta}}_m$ imply that

$$\frac{1}{N}|\sum_{i=1}^{k}\sum_{j=1}^{n_i}\{h^{\star}_{i,\ell}(\mathbf{x}_{ij},\widehat{\boldsymbol{\beta}}_m)-h^{\star}_{i,\ell}(\mathbf{x}_{ij},\boldsymbol{\beta}_m)\}|=\mathbf{o}_p(N^{-\frac{1}{2}})$$

Finally, from **S3** and the continuity of $\ddot{\rho}(\mathbf{a})$, the result holds using a Taylor expansion of order one of $\dot{\rho}(\mathbf{a})$. $\square$

We will now derive the Bahadur expansions given in 3.1. For that purpose, we need to obtain, as in Cui, He and Ng (2003) some identities satisfied by the common direction estimators.

Using the Lagrange multiplier method, we have that $\widehat{\boldsymbol{\beta}}_1$ maximises

$$G_1(\mathbf{a},\mu_1)=\rho_{N}(\mathbf{a})-\mu_1(\mathbf{a}^{\mathrm{T}}\mathbf{a}-1),$$

where $\mu_1\in\mathbb{R}$. Hence, differentiating G_1 respect to $\mathbf{a}$, we get that $\dot{\rho}_N(\widehat{\boldsymbol{\beta}}_1)=2\mu_1\widehat{\boldsymbol{\beta}}_1$, and so, using **S3**, the fact that $N^{1/2}(\tau_{iN}-\tau_i)\to 0$ and the equality

$$\begin{aligned}
\dot{\rho}_{N}(\mathbf{a}) &= \sum_{i=1}^{k}\tau_{iN}\dot{\varsigma}_{i,n_i}(\mathbf{a})=\sum_{i=1}^{k}\tau_{iN}\dot{\varsigma}_i(\mathbf{a})+\sum_{i=1}^{k}\tau_{iN}\frac{1}{n_i}\sum_{j=1}^{n_i}h^{\star}_i(\mathbf{x}_{ij},\mathbf{a})+\mathbf{o}_p(N^{-1/2})\\
&= \sum_{i=1}^{k}(\tau_{iN}-\tau_i)\dot{\varsigma}_i(\mathbf{a})+\sum_{i=1}^{k}\tau_i\dot{\varsigma}_i(\mathbf{a})+\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}h^{\star}_i(\mathbf{x}_{ij},\mathbf{a})+\mathbf{o}_p(N^{-1/2})\\
&= \dot{\rho}(\mathbf{a})+\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}h^{\star}_i(\mathbf{x}_{ij},\mathbf{a})+\mathbf{o}_p(N^{-1/2}) \qquad (A.4)
\end{aligned}$$

we obtain

$$\dot{\rho}(\widehat{\boldsymbol{\beta}}_1)+N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}h^{\star}_i(\mathbf{x}_{ij},\widehat{\boldsymbol{\beta}}_1) \;=\; 2\mu_1\widehat{\boldsymbol{\beta}}_1+\mathbf{o}_p\,(N^{-\frac{1}{2}}). \qquad (A.5)$$

Let, for $1\le m\le p$, $\widehat{\mathbf{P}}_{m+1}=\mathbf{I}_p-\sum_{j=1}^{m}\widehat{\boldsymbol{\beta}}_j\widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}$ be the projection matrix over the linear space orthogonal to that spaned by $\widehat{\boldsymbol{\beta}}_1,\ldots,\widehat{\boldsymbol{\beta}}_m$. Then, we have that $\widehat{\mathbf{P}}_2\widehat{\boldsymbol{\beta}}_1=0$ and so, we get

$$\widehat{\mathbf{P}}_2\left(\dot{\rho}(\widehat{\boldsymbol{\beta}}_1)+N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}h^{\star}_i(\mathbf{x}_{ij},\widehat{\boldsymbol{\beta}}_1)\right) \;=\; \mathbf{o}_p\,(N^{-\frac{1}{2}}).$$

Similarly, we have that $\widehat{\boldsymbol{\beta}}_m$ maximises

$$G_m(\mathbf{a},\mu_1,\ldots,\mu_m)=\rho_{N}(\mathbf{a})-\sum_{j=1}^{m-1}\mu_j\widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}\mathbf{a}-\mu_m(\mathbf{a}^{\mathrm{T}}\mathbf{a}-1),$$

for $m = 1, \dots q$ which implies that $\dot\rho_N(\widehat{\boldsymbol\beta}_m) = \sum_{j=1}^{m-1} \mu_j \widehat{\boldsymbol\beta}_j + 2\mu_m \widehat{\boldsymbol\beta}_m$. Therefore, using again **S3**, the fact that $N^{1/2}(\tau_{iN} - \tau_i) \to 0$, (A.4) and that $\widehat{\mathbf{P}}_{m+1}\widehat{\boldsymbol\beta}_j = 0$, $1 \le j \le m$, we obtain that

$$\widehat{\mathbf{P}}_{m+1}\left(\dot\rho(\widehat{\boldsymbol\beta}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij}, \widehat{\boldsymbol\beta}_m)\right) \;=\; \mathbf{o}_p\,(N^{-\frac{1}{2}}). \tag{A.6}$$

The equation (A.6) has its asymptotic version given in the next Lemma.

Lemma A.4. *Under the conditions of Theorem 4.2. the following equation holds.*

$$\mathbf{P}_{m+1}\dot\rho(\boldsymbol\beta_m) \;=\; 0 \tag{A.7}$$

Proof. Using Lemma A.3(ii) and the consistency of $\widehat{\boldsymbol\beta}_m$, we get that

$$N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij}, \widehat{\boldsymbol\beta}_m) = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij}, \boldsymbol\beta_m) + \mathbf{o}_p(1)$$

Replacing in (A.6), we obtain that

$$\widehat{\mathbf{P}}_{m+1}\left[\dot\rho(\widehat{\boldsymbol\beta}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij}, \boldsymbol\beta_m)\}\right] + \mathbf{o}_p(1) = \; 0\;.$$

Using the law of large numbers, the continuity of $\dot\rho$ and the consistency of $\widehat{\boldsymbol\beta}_m$, we obtain (A.7). $\square$

Moreover, we have the following relation between the estimators $\widehat{\boldsymbol\beta}_m$ and the projection matrix.

Lemma A.5. *Under conditions of Theorem 3.2. we have that, for all* $\mathbf{b} \in \mathbb{R}^p$

$$(\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{b} \;=\; -\sum_{j=1}^{m}(\;\boldsymbol\beta_j^{\mathrm{T}}\mathbf{b}\mathbf{I}_p \;+\; \boldsymbol\beta_j\mathbf{b}^{\mathrm{T}}\;)(\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_j) + \mathbf{O}_p(\;\|\mathbf{b}\|\sum_{i=1}^{m}\|\widehat{\boldsymbol\beta}_i - \boldsymbol\beta_i\|^2\;).$$

Proof. Indeed,

$$\sum_{j=1}^{m}(\;\boldsymbol\beta_j^{\mathrm{T}}\mathbf{b}\mathbf{I}_p + \boldsymbol\beta_j\mathbf{b}^{\mathrm{T}}\;)(\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_j) = \sum_{j=1}^{m}\boldsymbol\beta_j^{\mathrm{T}}\mathbf{b}\widehat{\boldsymbol\beta}_j - \sum_{j=1}^{m}\boldsymbol\beta_j^{\mathrm{T}}\mathbf{b}\boldsymbol\beta_j + \sum_{j=1}^{m}\boldsymbol\beta_j\mathbf{b}^{\mathrm{T}}\widehat{\boldsymbol\beta}_j - \sum_{j=1}^{m}\boldsymbol\beta_j\mathbf{b}^{\mathrm{T}}\boldsymbol\beta_j =$$

$$\sum_{j=1}^{m}\langle\boldsymbol\beta_j, \mathbf{b}\rangle\widehat{\boldsymbol\beta}_j + \sum_{j=1}^{m}\langle\widehat{\boldsymbol\beta}_j, \mathbf{b}\rangle\boldsymbol\beta_j - \sum_{j=1}^{m}\langle\boldsymbol\beta_j, \mathbf{b}\rangle\boldsymbol\beta_j - \sum_{j=1}^{m}\langle\boldsymbol\beta_j, \mathbf{b}\rangle\boldsymbol\beta_j.$$

Therefore,

$$\sum_{j=1}^{m}(\;\boldsymbol\beta_j^{\mathrm{T}}\mathbf{b}\mathbf{I}_p + \boldsymbol\beta_j\mathbf{b}^{\mathrm{T}}\;)(\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_j) = \sum_{j=1}^{m}\langle\boldsymbol\beta_j, \mathbf{b}\rangle\widehat{\boldsymbol\beta}_j + \sum_{j=1}^{m}\langle\widehat{\boldsymbol\beta}_j, \mathbf{b}\rangle\boldsymbol\beta_j - 2\sum_{j=1}^{m}\langle\boldsymbol\beta_j, \mathbf{b}\rangle\boldsymbol\beta_j.$$

On the other hand,

$$(\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{b} \;=\; \sum_{i=1}^{m} \boldsymbol{\beta}_i \boldsymbol{\beta}_i^{\mathrm{T}} \mathbf{b} - \sum_{i=1}^{m} \widehat{\boldsymbol{\beta}}_i \widehat{\boldsymbol{\beta}}_i^{\mathrm{T}} \mathbf{b} \;=\; \sum_{i=1}^{m} \langle \boldsymbol{\beta}_i, \mathbf{b} \rangle \boldsymbol{\beta}_i - \sum_{i=1}^{m} \langle \widehat{\boldsymbol{\beta}}_i, \mathbf{b} \rangle \widehat{\boldsymbol{\beta}}_i.$$

Adding the last two equations and using the proprieties of the inner product, we get

$$\sum_{j=1}^{m} (\, \boldsymbol{\beta}_j^{\mathrm{T}} \mathbf{b} \mathbf{I}_p + \boldsymbol{\beta}_j \mathbf{b}^{\mathrm{T}} \,)(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) + (\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{b} = \sum_{j=1}^{m} \langle \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j, \mathbf{b} \rangle (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j).$$

The Cauchy- Schwartz inequality implies that

$$\begin{aligned} \| \sum_{j=1}^{m} \langle \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j, \mathbf{b} \rangle (\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j) \| \quad &\leq \quad \sum_{j=1}^{m} |\langle \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j, \mathbf{b} \rangle| \|(\boldsymbol{\beta}_j - \widehat{\boldsymbol{\beta}}_j)\| \\ &\leq \quad \sum_{j=1}^{m} \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|^2 \|\mathbf{b}\|, \end{aligned}$$

and so, we get the desired result.□

The Lemma A.6 gives the key to obtain the equality given in (3.1).

Lemma A.6. *Under conditions of Theorem 3.2, we have that*

$$\begin{aligned} &\left\{ \mathbf{P}_{m+1} \ddot{\rho}(\boldsymbol{\beta}_m) - \; \boldsymbol{\beta}_m^{\mathrm{T}} \dot{\rho}(\boldsymbol{\beta}_m) \mathbf{I}_p - \boldsymbol{\beta}_m \dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}} \right\} (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) + \mathbf{o}_p(\, \|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\,\|) \; = \\ &\sum_{j=1}^{m-1} \left\{ \; \boldsymbol{\beta}_j^{\mathrm{T}} \, \dot{\rho}(\boldsymbol{\beta}_m) \, \mathbf{I}_p \;\; + \;\; \boldsymbol{\beta}_j \, \dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}} \; \right\} \, (\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) \; + \; N^{-\frac{1}{2}} \, \mathbf{P}_{m+1} \, \mathbf{u}_m \\ &+ \mathbf{o}_p \, (\; \sum_{i=1}^{m-1} \|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\| \;) + \mathbf{o}_p \, (N^{-\frac{1}{2}}) \end{aligned}$$

Proof. Let

$$L = \mathbf{P}_{m+1} \ddot{\rho}(\boldsymbol{\beta}_m)(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) - \boldsymbol{\beta}_m^{\mathrm{T}} \dot{\rho}(\boldsymbol{\beta}_m)(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) - \boldsymbol{\beta}_m \dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}} (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m).$$

Adding and subtracting $\widehat{\mathbf{P}}_{m+1} \, \ddot{\rho}(\boldsymbol{\beta}_m) \, (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m)$ in L we have that

$$\begin{aligned} L \;\; &= \;\; (\mathbf{P}_{m+1} - \widehat{\mathbf{P}}_{m+1}) \, \ddot{\rho}(\boldsymbol{\beta}_m) \, (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) \; - \; \boldsymbol{\beta}_m^{\mathrm{T}} \, \dot{\rho}(\boldsymbol{\beta}_m) \, (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) \\ &- \;\; \boldsymbol{\beta}_m \, \dot{\rho}^{\mathrm{T}}(\boldsymbol{\beta}_m) \, (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) + \widehat{\mathbf{P}}_{m+1} \, \ddot{\rho}(\boldsymbol{\beta}_m) \, (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m). \end{aligned}$$

Since $(\mathbf{P}_{m+1} - \widehat{\mathbf{P}}_{m+1}) = \mathbf{o}_p(1)$ the first term of L is a $\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|)$.

Using Lemma A.3(ii), we obtain

$$\begin{aligned} L \;\; &= \;\; \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) - \boldsymbol{\beta}_m^{\mathrm{T}} \dot{\rho}(\boldsymbol{\beta}_m)(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) - \boldsymbol{\beta}_m \dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}} (\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) \\ &+ \;\; \widehat{\mathbf{P}}_{m+1} \Big[\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m) - \{ \dot{\rho}(\boldsymbol{\beta}_m) + N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m) \} \Big] \\ &+ \;\; \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) + \mathbf{o}_p(N^{-\frac{1}{2}}). \end{aligned}$$

Now, by (A.6) we get $\widehat{\mathbf{P}}_{m+1}\Big[\dot\rho(\widehat{\boldsymbol\beta}_m) + N^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij},\widehat{\boldsymbol\beta}_m)]\Big] = \mathbf{o}_p\,(N^{-\frac12})$ and so, reordering the terms, we obtain

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m\|)+\mathbf{o}_p(N^{-\frac12})-(\,\boldsymbol\beta_m^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p\;+\;\boldsymbol\beta_m\dot\rho(\boldsymbol\beta_m)^{\mathrm T}\,)(\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m)\\ &- \widehat{\mathbf{P}}_{m+1}\Big[\dot\rho(\boldsymbol\beta_m)+N^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)\Big].\end{aligned}$$

Remind that $\mathbf{u}_m = -N^{-\frac12}\sum_{i=1}^k\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij},\boldsymbol\beta_m)$, thus, we have that

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m\|)+\mathbf{o}_p(N^{-\frac12})-(\,\boldsymbol\beta_m^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p\;+\;\boldsymbol\beta_m\dot\rho(\boldsymbol\beta_m)^{\mathrm T}\,)(\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m)\\ &+ \widehat{\mathbf{P}}_{m+1}\mathbf{u}_mN^{-\frac12}-\widehat{\mathbf{P}}_{m+1}\dot\rho(\boldsymbol\beta_m).\end{aligned}$$

Then, adding and subtracting $\mathbf{P}_{m+1}\mathbf{u}_mN^{-\frac12}$ and $\mathbf{P}_{m+1}\dot\rho(\boldsymbol\beta_m)$ to the last equation and using (A.7) we obtain

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m\|)+\mathbf{o}_p(N^{-\frac12})-(\,\boldsymbol\beta_m^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p\;+\;\boldsymbol\beta_m\dot\rho(\boldsymbol\beta_m)^{\mathrm T}\,)(\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m)\\ &+ (\widehat{\mathbf{P}}_{m+1}-\mathbf{P}_{m+1})\mathbf{u}_mN^{-\frac12}+\mathbf{P}_{m+1}\mathbf{u}_mN^{-\frac12}-(\widehat{\mathbf{P}}_{m+1}-\mathbf{P}_{m+1})\dot\rho(\boldsymbol\beta_m).\end{aligned}$$

The central limit theorem entails that $\mathbf{u}_m$ converges in distribution to a random normal variable. Therefore, $\mathbf{u}_m$ is bounded in probability. Using that $(\widehat{\mathbf{P}}_{m+1}-\mathbf{P}_{m+1})=\mathbf{o}_p(1)$, we get that

$$(\widehat{\mathbf{P}}_{m+1}-\mathbf{P}_{m+1})\mathbf{u}_mN^{-\frac12}=\mathbf{o}_p(N^{-\frac12}).$$

Hence,

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m\|)+\mathbf{o}_p(N^{-\frac12})-(\,\boldsymbol\beta_m^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p\;+\;\boldsymbol\beta_m\dot\rho(\boldsymbol\beta_m)^{\mathrm T}\,)(\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m)\\ &+ \mathbf{P}_{m+1}\mathbf{u}_mN^{-\frac12}-(\widehat{\mathbf{P}}_{m+1}-\mathbf{P}_{m+1})\dot\rho(\boldsymbol\beta_m).\end{aligned}$$

Using Lemma A.5, we get that

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m\|)+\mathbf{o}_p(N^{-\frac12})-(\,\boldsymbol\beta_m^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p\;+\;\boldsymbol\beta_m\dot\rho(\boldsymbol\beta_m)^{\mathrm T}\,)(\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m)\\ &\quad+\mathbf{P}_{m+1}\mathbf{u}_mN^{-\frac12}+\sum_{j=1}^m(\,\boldsymbol\beta_j^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p\;+\;\boldsymbol\beta_j\dot\rho(\boldsymbol\beta_m)^{\mathrm T}\,)(\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j)+\mathbf{O}_p(\,\|\dot\rho(\boldsymbol\beta_m)\|\sum_{j=1}^m\|\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j\|^2\,).\end{aligned}$$

On the other hand, the continuity of $\dot\rho$ in $\mathcal{S}_p$ implies that $\dot\rho$ is bounded, therefore, using that $\|\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j\|=\mathbf{o}_p(1)$, we obtain that

$$\mathbf{O}_p(\,\|\dot\rho(\boldsymbol\beta_m)\|\sum_{j=1}^m\|\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j\|^2)=\mathbf{o}_p(\,\sum_{j=1}^m\|\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j\|)=\mathbf{o}_p(\,\sum_{j=1}^{m-1}\|\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j\|)+\mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m\|),$$

which entails that

$$\begin{aligned}&(\mathbf{P}_{m+1}\ddot\rho(\boldsymbol\beta_m)-\boldsymbol\beta_m^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p-\boldsymbol\beta_m\dot\rho(\boldsymbol\beta_m)^{\mathrm T})(\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m)+\mathbf{o}_p(\,\|\widehat{\boldsymbol\beta}_m-\boldsymbol\beta_m\,\|)\;=\\ &\sum_{j=1}^{m-1}(\,\boldsymbol\beta_j^{\mathrm T}\dot\rho(\boldsymbol\beta_m)\mathbf{I}_p\;+\;\boldsymbol\beta_j\dot\rho(\boldsymbol\beta_m)^{\mathrm T}\,)(\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j)+N^{-\frac12}\mathbf{P}_{m+1}\mathbf{u}_m+\mathbf{o}_p(\,\sum_{j=1}^{m-1}\|\widehat{\boldsymbol\beta}_j-\boldsymbol\beta_j\|\,)+\mathbf{o}_p(N^{-\frac12})\end{aligned}$$

concluding the proof. $\square$

Using the last Lemma we are going to prove the next equality

Lemma A.7. *Under conditions of Theorem 3.2.*

$$\begin{aligned}\mathbf{A}_m(\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)+\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|) &= \sum_{j=1}^{m-1}\mathbf{B}_{jm}(\widehat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j)+N^{-\frac{1}{2}}\mathbf{P}_{m+1}\mathbf{u}_m\\ &+ \mathbf{o}_p(\sum_{j=1}^{m-1}\|\widehat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_j\|)+\mathbf{o}_p(N^{-\frac{1}{2}}) \qquad \text{(A.8)}\end{aligned}$$

where

$$\mathbf{A}_m=\mathbf{P}_{m+1}\ddot{\rho}(\boldsymbol{\beta}_m)-\boldsymbol{\beta}_m^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\mathbf{I}_p-\sum_{i=1}^{m-1}\boldsymbol{\beta}_i^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\boldsymbol{\beta}_m\boldsymbol{\beta}_i^{\mathrm{T}}$$

$$\mathbf{B}_{jm}=\boldsymbol{\beta}_j^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\mathbf{I}_p+\boldsymbol{\beta}_j\dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}}$$

PROOF. Since $\mathbf{P}_{m+1}\dot{\rho}(\boldsymbol{\beta}_m)=0$, we have that

$$\dot{\rho}(\boldsymbol{\beta}_m) \;=\; \sum_{j=1}^{m}\langle\dot{\rho}(\boldsymbol{\beta}_m);\boldsymbol{\beta}_j\rangle\boldsymbol{\beta}_j. \qquad \text{(A.9)}$$

Then, using Theorem 3.1, it will be enough to show that

$$\boldsymbol{\beta}_m^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)=O_p(\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|^2).$$

Using that $\widehat{\boldsymbol{\beta}}_m$ y $\boldsymbol{\beta}_m\in\mathcal{S}_p$, we get easily that

$$\langle\boldsymbol{\beta}_m,\widehat{\boldsymbol{\beta}}_m\rangle=1-\frac{\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|^2}{2},$$

and so,

$$\begin{aligned}\boldsymbol{\beta}_m^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m) &= \langle\boldsymbol{\beta}_m,\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\rangle=\langle\boldsymbol{\beta}_m-\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\rangle+\langle\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\rangle\\ &= -\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|^2+1-\langle\boldsymbol{\beta}_m,\widehat{\boldsymbol{\beta}}_m\rangle\\ &= -\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|^2+1-(1-\frac{\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|^2}{2})\\ &= -\frac{\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|^2}{2}.\end{aligned}$$

Thus, $\boldsymbol{\beta}_m^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m)=O_p(\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|^2)=o_p(\|\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_m\|)$, which concludes the prove. $\square$

Lemma A.8. *Under conditions of Theorem 3.2. we have*

a) $\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1=\mathbf{O}_p(N^{-\frac{1}{2}})$

b) $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m = \mathbf{O}_p(N^{-\frac{1}{2}})$

PROOF. a) Using (A.8) with $m = 1$, we get $\mathbf{A}_1(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + \mathbf{o}_p(\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\,) = N^{-\frac{1}{2}}\,\mathbf{P}_2\,\mathbf{u}_1 + \mathbf{o}_p(\,N^{-\frac{1}{2}}\,)$ which implies that $(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = \mathbf{o}_p(\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\,) + \mathbf{A}_1^{-1}\,N^{-\frac{1}{2}}\,\mathbf{P}_2\,\mathbf{u}_1 + \mathbf{o}_p(\,N^{-\frac{1}{2}}\,)$. Since $\mathbf{P}_2\,\mathbf{u}_1$ converges in distiribution to a normal random variable, we have that $\mathbf{A}_1^{-1}\,N^{-\frac{1}{2}}\,\mathbf{P}_2\,\mathbf{u}_1 = \mathbf{O}_p(N^{-\frac{1}{2}})$, and so, we have that $N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = N^{\frac{1}{2}}\,\mathbf{o}_p(\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\,) + \mathbf{O}_p(1) + \mathbf{o}_p(1)$. Thus,

$$N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \;=\; N^{\frac{1}{2}}\mathbf{o}_p(\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\,) + \mathbf{O}_p(1). \tag{A.10}$$

Hence,

$$N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\left[\frac{(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|} - \frac{\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\,)}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|}\right] = \mathbf{O}_p(1).$$

Since $\left\|\frac{(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|} - \frac{\mathbf{o}_p(\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\,)}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|}\right\| \xrightarrow{p} 1$, we have that $N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\| = O_p(1)$ as desired.

b) Let us show that $N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\| = O_p(1)$, for all $2 \le j \le m-1$ entails that $N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\| = O_p(1)$. Indeed, from (A.8), we get

$$\begin{aligned} N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) &= N^{\frac{1}{2}}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) + \sum_{j=1}^{m-1}\mathbf{A}_m^{-1}\mathbf{B}_{jm}N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) \\ &+ \mathbf{A}_m^{-1}\mathbf{P}_{m+1}\mathbf{u}_m + N^{\frac{1}{2}}\mathbf{o}_p(\sum_{j=1}^{m-1}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|) + \mathbf{o}_p(1) \end{aligned}$$

Since $\mathbf{P}_{m+1}\mathbf{u}_m$ converges in distribution to a normal random variable and using the inductive assumption, we get

$$N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) \;=\; N^{\frac{1}{2}}\;\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) + \mathbf{O}_p(1)\;,$$

which is is analoguous to (A.10), and so, we get easily that $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m = \mathbf{O}_p(N^{-\frac{1}{2}})$ concluding the proof. □

The Bahadur representation for $\widehat{\boldsymbol{\beta}}_m$ given in (3.1) follows now easily using (A.8) and Lemma A.8.

Now we are going to obtain the expansion given in (3.2). Similar arguments to those considered in Lemma A.3, allow to show that

$$\frac{1}{n_i}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m) = \frac{1}{n_i}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m) + o_p(n_i^{-\frac{1}{2}})\;,$$

then

$$\widehat{\lambda}_{im} - \lambda_{im} \;=\; \varsigma_{i,n_i}(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i(\boldsymbol{\beta}_m) = \varsigma_{i,n_i}(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i(\widehat{\boldsymbol{\beta}}_m) + \varsigma_i(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i(\boldsymbol{\beta}_m)$$

$$
\begin{aligned}
&= \frac{1}{n_i}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij},\boldsymbol{\beta}_m) + o_p(n_i^{-\frac{1}{2}}) + \dot{\varsigma}_i(\boldsymbol{\beta}_m)^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) + o_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) \\
&= \frac{1}{n_i}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij},\boldsymbol{\beta}_m) + o_p(n_i^{-\frac{1}{2}}),
\end{aligned}
$$

and so, the proof is concluded. $\square$

References

[1] Boente, G. & Orellana, L. (2001). A Robust Approach to Common Principal Components. In *Statistics in Genetics and in the Environmental Sciences*, eds. L. T. Fernholz, S. Morgenthaler, and W. Stahel, pp. 117-147. Basel: Birkhauser.

[2] Boente, G., Pires, A. M. & Rodrigues, I. M. (2002). Influence Functions and Outlier Detection under the Common Principal Components Model: A Robust Approach. *Biometrika*, **89**, 861-875.

[3] Boente, G., Pires, A. M. & Rodrigues, I. M. (2006). General projection-pursuit estimators for the common principal components model: influence funtions and Monte Carlo study. *Journal of Multivariate Analysis*, **97**, 124-147.

[4] Croux, C. and Ruiz–Gazen, A. (2005). High Breakdown Estimators for Principal Components: the Projection–Pursuit Approach Revisited. *Journal of Multivariate Analysis*, **95**, 206-226.

[5] Cui, H., He, X. & Ng, K. W. (2003). Asymptotic Distribution of Principal Components Based on Robust Dispersions. *Biometrika*, **90**, 953-966.

[6] Flury, B. K. (1984). Common Principal Components in k Groups. *Journal of the American Statistical Association*, **79**, 892-898.

[7] Flury, B. K. (1988). *Common Principal Components and Related Multivariate Models*. New York: John Wiley.

[8] Li, G. and Chen, Z. (1985). Projection–Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo. *Journal of the American Statistical Association*, **80**, 759-766.

[9] Pires, A. M. & Branco, J. (2002). Partial Influence Functions. *Journal of Multivariate Analysis*, **83** , 458-468.

[10] van der Vaart, A. & Wellner, J. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. New York: Springer.