

# Sign tests for weak principal directions

DAVY PAINDAVEINE<sup>1,2</sup>, JULIEN REMY<sup>1</sup> and THOMAS VERDEBOUT<sup>1</sup>

<sup>1</sup>*ECARES and Département de Mathématique, Université libre de Bruxelles, Avenue F.D. Roosevelt, 50, ECARES, CP114/04, B-1050, Brussels, Belgium. E-mail: tverdebo@ulb.ac.be*

<sup>2</sup>*Toulouse School of Economics, Université Toulouse Capitole, Esplanade de l'Université, 31080 Toulouse Cedex 06, France*

We consider inference on the first principal direction of a  $p$ -variate elliptical distribution. We do so in challenging double asymptotic scenarios for which this direction eventually fails to be identifiable. In order to achieve robustness not only with respect to such weak identifiability but also with respect to heavy tails, we focus on sign-based statistical procedures, that is, on procedures that involve the observations only through their direction from the center of the distribution. We actually consider the generic problem of testing the null hypothesis that the first principal direction coincides with a given direction of  $\mathbb{R}^p$ . We first focus on weak identifiability setups involving single spikes (that is, involving spectra for which the smallest eigenvalue has multiplicity  $p - 1$ ). We show that, irrespective of the degree of weak identifiability, such setups offer local alternatives for which the corresponding sequence of statistical experiments converges in the Le Cam sense. Interestingly, the limiting experiments depend on the degree of weak identifiability. We exploit this convergence result to build optimal sign tests for the problem considered. In classical asymptotic scenarios where the spectrum is fixed, these tests are shown to be asymptotically equivalent to the sign-based likelihood ratio tests available in the literature. Unlike the latter, however, the proposed sign tests are robust to arbitrarily weak identifiability. We show that our tests meet the asymptotic level constraint irrespective of the structure of the spectrum, hence also in possibly multi-spike setups. We fully characterize the non-null asymptotic distributions of the corresponding test statistics under weak identifiability, which allows us to quantify the corresponding local asymptotic powers. Finally, Monte Carlo exercises are conducted to assess the finite-sample relevance of our asymptotic results and a real-data illustration is provided.

*Keywords:* Le Cam's asymptotic theory of statistical experiments; local asymptotic normality; principal component analysis; sign tests; weak identifiability

## 1. Introduction

Most classical methods in multivariate statistics are based on Gaussian maximum likelihood estimators of location and scatter, that is, on the sample mean and sample covariance matrix. Irrespective of the considered problem (location or scatter problems, multivariate regression problems, principal component analysis, canonical correlation analysis, etc.), these methods exhibit poor efficiency properties when Gaussian assumptions are violated, particularly when the underlying distributions have heavy tails. Moreover, the resulting procedures are very sensitive to possible outliers in the data.

To improve on this, many robust procedures were developed. In particular, multivariate sign methods, that is, methods that use the observations only through their direction from the center of the distribution, have become increasingly popular. For location problems, multivariate sign tests were considered in Randles [27], Möttönen and Oja [19], Hallin and Paindaveine [10] and

Paindaveine and Verdebout [25], whereas sign procedures for scatter or shape matrices were considered in Tyler [37], Dümbgen [4], Hallin and Paindaveine [11], Dürre, Vogel and Fried [7] and Dürre, Fried and Vogel [5], to cite only a few. PCA techniques based on multivariate signs (and on the companion concept of ranks) were studied in Hallin, Paindaveine and Verdebout [12], Taskinen, Koch and Oja [32], Hallin, Paindaveine and Verdebout [13] and Dürre, Tyler and Vogel [6]. Multivariate sign tests were also developed, for example, for testing i.i.d.-ness against serial dependence (see Paindaveine [23]), or for testing for multivariate independence (see Taskinen, Kankainen and Oja [31] and Taskinen, Oja and Randles [33]). Most references above actually focus on *spatial sign procedures*, that is, on procedures that are based on the signs  $\mathbf{U}_i := \mathbf{X}_i / \|\mathbf{X}_i\|$ ,  $i = 1, \dots, n$ , obtained by projecting the  $p$ -variate observations at hand onto the unit sphere  $\mathcal{S}^{p-1} := \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x} = 1\}$  of  $\mathbb{R}^p$  (sometimes, the projection is performed on standardized observations in order to achieve affine invariance). Since they discard the radii  $\|\mathbf{X}_i\|$ ,  $i = 1, \dots, n$ , spatial sign procedures can deal with arbitrarily heavy tails and are robust to observations that would be far from the center of the distribution. Furthermore, they are by nature well adapted to directional or axial data for which these radii are not observed; see, for example, Mardia and Jupp [18] or Ley and Verdebout [16]. For more details on spatial sign methods, we refer to the monograph Oja [21].

The present paper considers principal component analysis, or more precisely, inference on principal component directions. Consider the case where the observations  $\mathbf{X}_1, \dots, \mathbf{X}_n$  at hand form a random sample from a centered  $p$ -variate elliptical distribution, that is, a distribution whose characteristic function is of the form  $\mathbf{t} \mapsto \exp(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$  for some symmetric positive definite  $p \times p$  matrix  $\boldsymbol{\Sigma}$ . Assume that the ordered eigenvalues of  $\boldsymbol{\Sigma}$  satisfy  $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_p$ , so that the leading eigenvector  $\boldsymbol{\theta}_1$ , possibly unlike the other eigenvectors  $\boldsymbol{\theta}_j$ ,  $j = 2, \dots, p$ , is identifiable (as usual, identifiability is up to an unimportant sign). We will then throughout consider the problem of testing the null hypothesis  $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$  against the alternative hypothesis  $\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$ , where  $\boldsymbol{\theta}_0$  is a fixed unit  $p$ -vector. This testing problem has attracted much attention in the past decades, and the textbook procedure, namely the Anderson [1] Gaussian likelihood ratio test, has been extended in various directions. To mention only a few, Jolicoeur [15] considered a small-sample test, whereas Flury [8] proposed an extension to a larger number of eigenvectors. Tyler [34,35] robustified the Anderson [1] test to possible elliptical departures from multinormality (the original likelihood ratio test requires Gaussian assumptions). Schwartzman, Mascarenhas and Taylor [30] considered extensions to the case of Gaussian random matrices, and Hallin, Paindaveine and Verdebout [12] obtained Le Cam optimal tests for the problem considered.

All asymptotic tests above assume that the eigenvalues  $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_p$  are fixed, so that the eigenvector  $\boldsymbol{\theta}_1$  remains asymptotically identifiable. The null asymptotic distribution of the corresponding test statistics, however, may very poorly approximate their fixed- $n$  distribution when  $\lambda_1/\lambda_2$  is close to one. In the present work, we therefore consider general asymptotic scenarios that address this issue. More precisely, we allow for scatter values  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_n$  for which the corresponding ratio  $\lambda_{n1}/\lambda_{n2} (> 1)$  converges to one as  $n$  diverges to infinity. In such asymptotic scenarios, the leading eigenvector  $\boldsymbol{\theta}_{n1}$  remains identifiable for any  $n$  but is no longer identifiable in the limit. One then says that  $\boldsymbol{\theta}_{n1}$  is *weakly identifiable*. The distributional framework considered here formalizes situations that are often encountered in practice where two sample eigenvalues are close to each other so that inference about the corresponding eigenvectors is a priori difficult. Inference on weakly identified parameters has already been much considered

in the literature: see, e.g., Pötscher [26], Forchini and Hillier [9], Dufour [3] and Antoine and Lavergne [2]. Liu and Shao [17] and Zhu and Zhang [39] consider asymptotic inference under a total lack of identifiability. Recently, Paindaveine, Remy and Verdebout [24] considered Gaussian tests on weakly identified eigenvectors. While these tests can handle weak identifiability, they are based on sample covariance matrices, hence cannot deal with heavy tails and are very sensitive to possible outliers. In the present work, we tackle the same problem but develop spatial sign tests that not only inherit the robustness of sign procedures but also can deal with both heavy tails and weak identifiability.

To ensure that spatial signs are well-defined with probability one, we will restrict to elliptical distributions that do not attribute a positive probability mass to the symmetry center (the symmetry center will be assumed to be known and, without any loss of generality, to coincide with the origin of  $\mathbb{R}^p$  – extension to the unknown location case is straightforward, as we will explain in Section 7). Suitable spatial sign tests are to be determined in the image of the model by the projection onto the unit sphere  $S^{p-1}$ . Now, if the random  $p$ -vector  $\mathbf{X}$  is centered elliptical with scatter matrix  $\Sigma$ , then the corresponding spatial sign  $\mathbf{U} := \mathbf{X}/\|\mathbf{X}\|$  follows the *angular Gaussian distribution* with shape matrix  $\mathbf{V} := p\Sigma/\text{tr}(\Sigma)$ ; see Tyler [38]. Since  $\Sigma$  and  $\mathbf{V}$  share the same eigenvectors, the induced testing problem is still the problem of testing  $\mathcal{H}_0 : \theta_1 = \theta_0$  against  $\mathcal{H}_1 : \theta_1 \neq \theta_0$ , where  $\theta_1$  is the leading eigenvector of  $\mathbf{V}$ . Also, since eigenvalues of  $\Sigma$  and  $\mathbf{V}$  are equal up to a common positive factor, (weak) identifiability occurs for the original elliptical problem if and only if it does for the induced angular Gaussian problem. These considerations explain that identifying optimal spatial sign tests under weak identifiability for the elliptical problem should be done in the setup where one observes triangular arrays of spatial signs  $\mathbf{U}_{n1}, \dots, \mathbf{U}_{nn}$ ,  $n = 1, 2, \dots$ , randomly drawn from the angular Gaussian distribution with shape matrix  $\mathbf{V}_n$ .

Accordingly, we consider the problem of testing the null hypothesis  $\mathcal{H}_0 : \theta_{n1} = \theta_0$  in an angular Gaussian double asymptotic scenario for which  $\lambda_{n1}/\lambda_{n2} (> 1)$  may converge to one at an arbitrary rate, which provides weak identifiability for the leading principal direction  $\theta_{n1}$ . We first focus on sequences of single-spike shape matrices (characterized by spectra of the form  $\lambda_{n1} > \lambda_{n2} = \dots = \lambda_{np}$ ). We will show that, irrespective of the rate of convergence of  $\lambda_{n1}/\lambda_{n2}$  to one, there exist suitable local alternatives for which the corresponding sequence of statistical experiments converges in the Le Cam sense. Interestingly, the limiting experiment depends on the degree of weak identifiability. This paves the way, in this single-spike setup, to optimal testing for the problem considered. Quite nicely, we will actually build tests that are Le Cam optimal in single-spike setups while remaining valid (in the sense that they meet the asymptotic level constraint) under general, “multi-spike”, spectra. We will also fully characterize the non-null behavior of these tests under weak identifiability, which will allow us to extensively quantify their local asymptotic powers.

The outline of the paper is as follows. In Section 2, we introduce the notation, describe the sequence of angular Gaussian models to be considered, and derive the aforementioned results on limiting experiments. This is used to derive a sign test that can deal with weak identifiability and enjoys nice Le Cam optimality properties. However, (i) since this test involves nuisance parameters, it is an infeasible statistical procedure. Moreover, (ii) the optimal sign test requires a single-spike spectrum. In Section 3, we therefore derive a version of the optimal test that (i) is feasible and (ii) should be able to cope with multi-spike spectra. We show that, under the null

hypothesis, hence also under sequences of contiguous hypotheses, the infeasible and feasible tests are asymptotically equivalent, so that the latter inherits the optimality properties of the former. We also show that, in classical asymptotic scenarios where one stays away from weak identifiability, the proposed test is asymptotically equivalent to the likelihood ratio test from Tyler [38]. In Section 4, we investigate the asymptotic properties of the proposed sign test. We first show that, as anticipated above, this test asymptotically achieves the target null size even when the underlying scatter matrix does not have a single-spike structure. Then, for any degree of weak identifiability, we derive the asymptotic distribution of the proposed test statistic under suitable local alternatives. In Section 5, Monte Carlo exercises are conducted (a) to show that the proposed sign test, unlike its competitors, can deal with both heavy tails and weak identifiability and (b) to compare the finite-sample powers of the various tests. A real-data illustration is presented in Section 6. Finally, conclusions and final comments are provided in Section 7. All proofs are collected in the technical Appendices A–C.

## 2. Limits of angular Gaussian experiments

In this section, our objective is to derive the form of locally asymptotically optimal tests for  $\mathcal{H}_0 : \boldsymbol{\theta}_{n1} = \boldsymbol{\theta}_0$  in asymptotic scenarios under which  $\boldsymbol{\theta}_{n1}$  is weakly identifiable. To do so, we study sequences of angular Gaussian experiments indexed by shape matrices  $\mathbf{V}_n$  with eigenvalues  $\lambda_{n1} > \lambda_{n2} = \dots = \lambda_{np}$  satisfying  $\lambda_{n1}/\lambda_{n2} \rightarrow 1$ . Since the optimal tests we will obtain in this section are actually infeasible, we will then construct in Section 3 a practical sign test that achieves the same (null and non-null) asymptotic properties as the infeasible tests; the present section can therefore be seen as a stepping stone to the tests proposed in Section 3.

Consider a triangular array of observations  $\mathbf{X}_{ni}, i = 1, \dots, n, n = 1, 2, \dots$ , such that for any  $n$ , the random  $p$ -vectors  $\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}$  form a random sample from a centered elliptical distribution that does not attribute a positive probability mass to the origin of  $\mathbb{R}^p$ . For any  $n$ , the leading eigenvector  $\boldsymbol{\theta}_{n1}$  of the corresponding scatter matrix  $\boldsymbol{\Sigma}_n$  is assumed to be well identified, up to a sign. In this general setup, we aim at designing suitable sign tests for the problem of testing the null hypothesis  $\mathcal{H}_0 : \boldsymbol{\theta}_{n1} = \boldsymbol{\theta}_0$  against the alternative hypothesis  $\mathcal{H}_1 : \boldsymbol{\theta}_{n1} \neq \boldsymbol{\theta}_0$ , where  $\boldsymbol{\theta}_0$  is a fixed unit  $p$ -vector. As explained in the Introduction, such tests should be determined in the image of the model by the projection of the model onto the unit sphere  $\mathcal{S}^{p-1}$ . For any  $n$ , the projected observations  $\mathbf{U}_{n1}, \dots, \mathbf{U}_{nn}$ , where  $\mathbf{U}_{ni} = \mathbf{X}_{ni}/\|\mathbf{X}_{ni}\|$  is the spatial sign of  $\mathbf{X}_{ni}$ , form a random sample from the  $p$ -variate angular Gaussian distribution with shape matrix  $\mathbf{V}_n = p\boldsymbol{\Sigma}_n/\text{tr}(\boldsymbol{\Sigma}_n)$  (shape matrices throughout are normalized to have trace  $p$ ), so that the density of  $\mathbf{U}_{n1}$  with respect to the surface area measure on  $\mathcal{S}^{p-1}$  is

$$\mathbf{u} (\in \mathcal{S}^{p-1}) \mapsto \frac{\Gamma(\frac{p}{2})}{2\pi^{p/2}(\det \mathbf{V}_n)^{1/2}} (\mathbf{u}'\mathbf{V}_n^{-1}\mathbf{u})^{-p/2}, \tag{1}$$

where  $\Gamma(\cdot)$  is Euler’s gamma function; see Tyler [38]. We will denote the corresponding hypothesis as  $\mathbf{P}_{\mathbf{V}_n}^{(n)}$  (at places,  $\mathbf{P}_{\mathbf{V}_n}^{(n)}$  will also denote the distribution of a random sample of size  $n$  from the angular Gaussian distribution with shape matrix  $\mathbf{V}_n$ ). In this angular Gaussian framework, the aforementioned elliptical testing problem induces the problem of testing the null hypothesis

$\mathcal{H}_0 : \boldsymbol{\theta}_{n1} = \boldsymbol{\theta}_0$  against the alternative hypothesis  $\mathcal{H}_1 : \boldsymbol{\theta}_{n1} \neq \boldsymbol{\theta}_0$ , where  $\boldsymbol{\theta}_{n1}$  is the leading eigenvector of  $\mathbf{V}_n$  (recall that  $\mathbf{V}_n$  and  $\boldsymbol{\Sigma}_n$  share the same eigenvectors). Throughout,  $\lambda_{nj}$  and  $\boldsymbol{\theta}_{nj}$ ,  $j = 1, \dots, p$ , will refer to the (ordered) eigenvalues and eigenvectors of  $\mathbf{V}_n$ . In the original elliptical problem, weak identifiability of  $\boldsymbol{\theta}_{n1}$ , meaning that  $\boldsymbol{\theta}_{n1}$  is identified for any fixed  $n$  but is not in the limit as  $n \rightarrow \infty$ , clearly occurs if and only if  $\lambda_{n1}/\lambda_{n2} \rightarrow 1$ .

The goal of this section is to determine optimal sign tests, under possibly weak identifiability of  $\boldsymbol{\theta}_{n1}$ , in the particular case of single-spike spectra, that is, in situations where the smallest eigenvalue has multiplicity  $p - 1$ . These tests will result from the asymptotic study of angular Gaussian log-likelihood ratios in Theorem 2.1 below. We will consider sequences of null shape matrices of the form

$$\mathbf{V}_{0n} := \left(1 - \frac{\delta_n \xi}{p}\right) \mathbf{I}_p + \delta_n \xi \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0, \tag{2}$$

where  $\xi > 0$  is a *locality parameter* and  $\delta_n$  is a bounded positive sequence that may be  $o(1)$ ; throughout, we tacitly assume that  $\xi$  and  $\delta_n$  are chosen so that  $\mathbf{V}_{0n}$  is positive definite. It is straightforward to check that  $\mathbf{V}_{0n}$  indeed has a single spike: the largest eigenvalue is  $\lambda_{n1} = 1 + (p - 1)\delta_n \xi/p$ , with multiplicity one and corresponding eigenvector  $\boldsymbol{\theta}_0$ , whereas the remaining eigenvalues are  $\lambda_{n2} = \dots = \lambda_{np} = 1 - \delta_n \xi/p$ , with an eigenspace that is the orthogonal complement to  $\boldsymbol{\theta}_0$ . In this setup, weak identifiability of  $\boldsymbol{\theta}_{n1}$  occurs if and only if  $\delta_n$  is  $o(1)$ .

To discuss optimality issues, we consider local alternatives associated with perturbations  $\boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_n$  of  $\boldsymbol{\theta}_0$ , where  $(v_n)$  is a positive sequence and  $(\boldsymbol{\tau}_n)$  is a bounded sequence in  $\mathbb{R}^p$  such that  $\boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_n \in \mathcal{S}^{p-1}$  for any  $n$ . It is easy to show that the latter condition entails that  $v_n$  and  $\boldsymbol{\tau}_n$  must satisfy

$$\boldsymbol{\theta}'_0 \boldsymbol{\tau}_n = -\frac{v_n}{2} \|\boldsymbol{\tau}_n\|^2 \tag{3}$$

for any  $n$ . The resulting sequence of alternatives is then associated with the single-spike shape matrices

$$\mathbf{V}_{1n} := \left(1 - \frac{\delta_n \xi}{p}\right) \mathbf{I}_p + \delta_n \xi (\boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_n)(\boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_n)'. \tag{4}$$

Our construction of optimal sign tests requires studying the asymptotic behavior, under  $\mathbf{P}_{\mathbf{V}_{0n}}^{(n)}$ , of the log-likelihood ratios  $\Lambda_n := \log(d\mathbf{P}_{\mathbf{V}_{1n}}^{(n)}/d\mathbf{P}_{\mathbf{V}_{0n}}^{(n)})$ . To do so, let

$$\gamma_n := \frac{p\delta_n \xi}{p + (p - 1)\delta_n \xi} \quad \text{and} \quad \mathbf{S}_n(\mathbf{V}) := \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{V}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \mathbf{V}^{-1/2}}{\|\mathbf{V}^{-1/2} \mathbf{U}_{ni}\|^2}, \tag{5}$$

where  $\mathbf{V}^{-1/2}$  is the symmetric positive definite square root of  $\mathbf{V}^{-1}$ . Note that the sequence  $(\gamma_n)$  is  $O(\delta_n)$ . We then have the following result.

**Theorem 2.1.** *Fix  $\boldsymbol{\theta}_0 \in \mathcal{S}^{p-1}$ . Let  $(v_n)$  be a positive sequence and  $(\boldsymbol{\tau}_n)$  be a bounded sequence in  $\mathbb{R}^p$  such that  $\boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_n \in \mathcal{S}^{p-1}$  for any  $n$ . Let  $\mathbf{V}_{0n}$  be as in (2). Then, as  $n \rightarrow \infty$ , under  $\mathbf{P}_{\mathbf{V}_{0n}}^{(n)}$ , we have the following:*

(i) if  $\delta_n \equiv 1$ , then, for  $v_n = 1/(\sqrt{n}\gamma_n)$ ,

$$\Lambda_n = \boldsymbol{\tau}'_n \boldsymbol{\Delta}_{\boldsymbol{\theta}_0}^{(i)} - \frac{1}{2} \boldsymbol{\tau}'_n \boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}^{(i)} \boldsymbol{\tau}_n + o_P(1) \tag{6}$$

as  $n \rightarrow \infty$ , where we let

$$\boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}^{(i)} := \frac{p(p + (p - 1)\xi)}{(p + 2)(p - \xi)} (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0),$$

and where

$$\boldsymbol{\Delta}_{\boldsymbol{\theta}_0}^{(i)} := p \frac{\sqrt{p + (p - 1)\xi}}{\sqrt{p - \xi}} (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \sqrt{n} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \boldsymbol{\theta}_0$$

is asymptotically normal with mean zero and covariance matrix  $\boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}^{(i)}$ ;

(ii) if  $\delta_n$  is  $o(1)$  with  $\sqrt{n}\delta_n \rightarrow \infty$ , then, for  $v_n = 1/(\sqrt{n}\gamma_n)$ ,

$$\Lambda_n = \boldsymbol{\tau}'_n \boldsymbol{\Delta}_{\boldsymbol{\theta}_0}^{(ii)} - \frac{1}{2} \boldsymbol{\tau}'_n \boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}^{(ii)} \boldsymbol{\tau}_n + o_P(1) \tag{7}$$

as  $n \rightarrow \infty$ , where we let

$$\boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}^{(ii)} := \frac{p}{p + 2} (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \tag{8}$$

and where

$$\boldsymbol{\Delta}_{\boldsymbol{\theta}_0}^{(ii)} := p (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \sqrt{n} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \boldsymbol{\theta}_0 \tag{9}$$

is asymptotically normal with mean zero and covariance matrix  $\boldsymbol{\Gamma}_{\boldsymbol{\theta}_0}$ ;

(iii) if  $\delta_n = 1/\sqrt{n}$ , then, for  $v_n = 1/(\sqrt{n}\gamma_n)$  (or equivalently  $v_n \equiv 1/\xi$ ),

$$\Lambda_n = \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\theta}_0 + \frac{1}{2\xi} \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\tau}_n - \frac{p}{2(p + 2)} \left( \|\boldsymbol{\tau}_n\|^2 - \frac{1}{4\xi^2} \|\boldsymbol{\tau}_n\|^4 \right) + o_P(1), \tag{10}$$

where

$$\boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} := p \sqrt{n} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right)$$

is such that  $\text{vec}(\boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)})$  is asymptotically normal with mean zero and covariance matrix  $(p/(p + 2))(\mathbf{I}_{p^2} + \mathbf{K}_p + \mathbf{J}_p) - \mathbf{J}_p$  (here,  $\mathbf{K}_p := \sum_{i,j=1}^p (\mathbf{e}_i \mathbf{e}'_j) \otimes (\mathbf{e}_j \mathbf{e}'_i)$  and  $\mathbf{J}_p := \sum_{i,j=1}^p (\mathbf{e}_i \mathbf{e}'_j) \otimes (\mathbf{e}_j \mathbf{e}'_j) = (\text{vec } \mathbf{I}_p)(\text{vec } \mathbf{I}_p)'$ , where  $\mathbf{e}_\ell$  is the  $\ell$ th vector of the canonical basis of  $\mathbb{R}^p$  and  $\mathbf{A} \otimes \mathbf{B}$  is the Kronecker product between the matrices  $\mathbf{A}$  and  $\mathbf{B}$ );

(iv) if  $\sqrt{n}\delta_n \rightarrow 0$ , then, even for  $v_n \equiv 1/\xi$ , we have  $\Lambda_n = o_P(1)$ .

Theorem 2.1 shows that the asymptotic behavior of  $\Lambda_n$  crucially depends on the sequence  $(\delta_n)$  and identifies four different regimes. A similar phenomenon has been obtained in Tyler [36]

when investigating the limiting behavior of eigenvalues of scatter estimators (in particular, the largest eigenvalue of the scatter estimators considered in Tyler [36] shows a limiting behavior that depends on  $\delta_n$  and, parallel to what we have in the present work,  $\delta_n = 1/\sqrt{n}$  also turns out to be an important threshold there). In the “classical” regime (i) where  $\delta_n \equiv 1$ , standard perturbations with  $v_n \sim 1/\sqrt{n}$  provide a sequence of experiments that is locally asymptotically normal (LAN), with central sequence  $\Delta_{\theta_0}^{(i)}$  and Fisher information  $\Gamma_{\theta_0}^{(i)}$ . In such a LAN setup, the locally asymptotically maximin test  $\phi^{(n)}$  (see, e.g., Section 5.2.3 from Ley and Verdebout [16] for the concept of maximin tests) for  $\mathcal{H}_0 : \theta_{n1} = \theta_0$  against  $\mathcal{H}_1 : \theta_{n1} \neq \theta_0$  rejects the null hypothesis at asymptotic level  $\alpha$  when

$$T_n(\mathbf{V}_{0n}) := (\Delta_{\theta_0}^{(i)})' (\Gamma_{\theta_0}^{(i)})^{-1} \Delta_{\theta_0}^{(i)} = np(p+2) \|(\mathbf{I}_p - \theta_0 \theta_0') \mathbf{S}_n(\mathbf{V}_{0n}) \theta_0\|^2 > \chi_{p-1, 1-\alpha}^2, \quad (11)$$

where  $\mathbf{A}^-$  stands for the Moore–Penrose inverse of  $\mathbf{A}$  and where  $\chi_{\ell, 1-\alpha}^2$  denotes the upper  $\alpha$ -quantile of the chi-square distribution with  $\ell$  degrees of freedom. In regime (ii), perturbations with  $v_n \sim 1/(\sqrt{n}\delta_n)$  – that is, perturbations that are more severe than in the standard regime (i) – make the sequence of experiments LAN, here with the central sequence  $\Delta_{\theta_0}^{(ii)}$  in (9) and the Fisher information matrix  $\Gamma_{\theta_0}^{(ii)}$  in (8). Since

$$(\Delta_{\theta_0}^{(ii)})' (\Gamma_{\theta_0}^{(ii)})^{-1} \Delta_{\theta_0}^{(ii)} = T_n(\mathbf{V}_{0n}),$$

the test  $\phi^{(n)}$  is still locally asymptotically maximin in regime (ii).

The situation in regime (iii) is quite different. While the sequence of experiments there is not LAN nor LAMN (locally asymptotically mixed normal), it still converges in the Le Cam sense. It is easy to check that, in this regime,

$$\tau_n' \Upsilon_{\theta_0}^{(n)} \theta_0 + \frac{1}{2\xi} \tau_n' \Upsilon_{\theta_0}^{(n)} \tau_n$$

is asymptotically normal with mean zero and variance

$$\frac{p}{p+2} \left( \|\tau_n\|^2 - \frac{1}{4\xi^2} \|\tau_n\|^4 \right),$$

so that the Le Cam first lemma entails that the sequences of hypotheses  $P_{\mathbf{V}_{0n}}^{(n)}$  and  $P_{\mathbf{V}_{1n}}^{(n)}$  are mutually contiguous in regime (iii), too. As we will show in the next section, the test  $\phi^{(n)}$  shows non-trivial asymptotic powers under these contiguous alternatives. Finally, in regime (iv), Theorem 2.1 shows that no test can discriminate between the null hypothesis and the alternatives associated with  $v_n \sim 1$ , which are the most severe ones that can be considered.

### 3. The proposed sign test

The test  $\phi^{(n)}$  from the previous section enjoys nice optimality properties. However, (i) it is unfortunately infeasible (its test statistic  $T_n(\mathbf{V}_{0n})$  indeed involves the population shape matrix  $\mathbf{V}_{0n}$ ,

which is of course unknown in practice); moreover, (ii) the test  $\phi^{(n)}$  will in principle meet the asymptotic level constraint only under the, quite restrictive, single-spike shape structure in (2). In this section, we therefore construct a version  $\tilde{\phi}^{(n)}$  of  $\phi^{(n)}$  that (i) is feasible and that (ii) will be able to cope with more general, multi-spike, shape structures.

To do so, let  $(\mathbf{V}_{0n})$  be a sequence of shape matrices associated with the null hypothesis. In other words, we assume that, for any  $n$ , the shape  $\mathbf{V}_{0n}$  admits the spectral decomposition

$$\mathbf{V}_{0n} = \lambda_{n1}\boldsymbol{\theta}_0\boldsymbol{\theta}'_0 + \sum_{j=2}^p \lambda_{nj}\boldsymbol{\theta}_{nj}\boldsymbol{\theta}'_{nj}, \tag{12}$$

where the  $\boldsymbol{\theta}_{nj}$ 's form an orthonormal basis of the orthogonal complement to  $\boldsymbol{\theta}_0$  and where  $\lambda_{n1} > \lambda_{n2} \geq \dots \geq \lambda_{np}$ . In particular, the  $p - 1$  smallest eigenvalues here do not need to be equal, so that the number of ‘‘spikes’’ may be arbitrary. With this notation, it is clear that estimating  $\mathbf{V}_{0n}$  requires estimating the eigenvalues  $\lambda_{n1}, \dots, \lambda_{np}$  and the eigenvectors  $\boldsymbol{\theta}_{n2}, \dots, \boldsymbol{\theta}_{np}$ . To do so, we consider the Tyler [37] M-estimator, that is defined as the shape matrix satisfying  $\mathbf{S}_n(\hat{\mathbf{V}}_n) = (1/p)\mathbf{I}_p$ ; see (5). Note that  $T_n(\hat{\mathbf{V}}_n) = 0$  almost surely, so that  $\hat{\mathbf{V}}_n$  cannot be used to estimate  $\mathbf{V}_{0n}$  directly. Decompose then the M-estimator  $\hat{\mathbf{V}}_n$  into

$$\hat{\mathbf{V}}_n = \sum_{j=1}^p \hat{\lambda}_{nj}\hat{\boldsymbol{\theta}}_{nj}\hat{\boldsymbol{\theta}}'_{nj}. \tag{13}$$

The eigenvalues  $\hat{\lambda}_{n1}, \dots, \hat{\lambda}_{np}$  of Tyler’s M-estimator provide estimates of the eigenvalues in (12). Later asymptotic results, however, will require that the estimators of the eigenvectors  $\boldsymbol{\theta}_{nj}$  are orthogonal to the null value  $\boldsymbol{\theta}_0$  of the first eigenvector  $\boldsymbol{\theta}_{n1}$ , a constraint that the eigenvectors  $\hat{\boldsymbol{\theta}}_{nj}$  do not meet in general. To correct for this, we will rather use the estimators  $\tilde{\boldsymbol{\theta}}_{nj}, j = 2, \dots, p$ , resulting from a Gram–Schmidt orthogonalization of  $\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_{n2}, \dots, \hat{\boldsymbol{\theta}}_{np}$ . In other words,  $\tilde{\boldsymbol{\theta}}_{nj}$  is defined recursively through

$$\tilde{\boldsymbol{\theta}}_{nj} := \frac{(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0 - \sum_{k=2}^{j-1} \tilde{\boldsymbol{\theta}}_{nk}\tilde{\boldsymbol{\theta}}'_{nk})\hat{\boldsymbol{\theta}}_{nj}}{\|(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0 - \sum_{k=2}^{j-1} \tilde{\boldsymbol{\theta}}_{nk}\tilde{\boldsymbol{\theta}}'_{nk})\hat{\boldsymbol{\theta}}_{nj}\|}, \quad j = 2, \dots, p, \tag{14}$$

with summation over an empty collection of indices being equal to zero.

In the rest of the paper,  $\tilde{\mathbf{V}}_{0n}$  will denote the estimator of  $\mathbf{V}_{0n}$  obtained by substituting in (12) the Tyler eigenvalues  $\hat{\lambda}_{nj}$  and eigenvectors  $\tilde{\boldsymbol{\theta}}_{nj}$  for the  $\lambda_{nj}$ 's and  $\boldsymbol{\theta}_{nj}$ 's. Since Tyler’s M-estimator is normalized to have trace  $p$ , the estimator  $\tilde{\mathbf{V}}_{0n}$  also has trace  $p$ , hence is a shape matrix. Further note that if a single-spiked model as in Section 2 is assumed, which is a common practice for large  $p$ , then  $\tilde{\mathbf{V}}_{0n}$  can be replaced by  $\hat{\lambda}_{n1}\boldsymbol{\theta}_0\boldsymbol{\theta}'_0 + ((p - 1)^{-1} \sum_{j=2}^p \hat{\lambda}_{nj})(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0)$ . Quite nicely, replacing  $\mathbf{V}_{0n}$  with  $\tilde{\mathbf{V}}_{0n}$  in the test statistic  $T_n(\mathbf{V}_{0n})$  of  $\phi^{(n)}$  has no asymptotic impact in probability under the null hypothesis. More precisely, we have the following result.

**Theorem 3.1.** *Let  $(\mathbf{V}_{0n})$  be a sequence of null shape matrices as in (12). Then, under  $P_{\mathbf{V}_{0n}}^{(n)}$ ,  $T_n(\tilde{\mathbf{V}}_{0n}) = T_n(\mathbf{V}_{0n}) + o_P(1)$  as  $n \rightarrow \infty$ .*

Based on this result, the sign test we propose in this paper is the test  $\tilde{\phi}^{(n)}$  that rejects the null hypothesis  $\mathcal{H}_0 : \theta_{n1} = \theta_0$  at asymptotic level  $\alpha$  whenever

$$T_n(\tilde{\mathbf{V}}_{0n}) = np(p + 2) \|(\mathbf{I}_p - \theta_0 \theta_0') \mathbf{S}_n(\tilde{\mathbf{V}}_{0n}) \theta_0\|^2 > \chi_{p-1, 1-\alpha}^2. \tag{15}$$

Unlike  $\phi^{(n)}$ , this new sign test is a feasible statistical procedure. An alternative test for the same problem is the likelihood ratio test,  $\phi_{\text{Ty1}}^{(n)}$  say, that rejects the null hypothesis at asymptotic level  $\alpha$  whenever

$$L_n := \frac{np}{p + 2} (\hat{\lambda}_{n1} \theta_0' \hat{\mathbf{V}}_n^{-1} \theta_0 + \hat{\lambda}_{n1}^{-1} \theta_0' \hat{\mathbf{V}}_n \theta_0 - 2) > \chi_{p-1, 1-\alpha}^2, \tag{16}$$

where  $\hat{\mathbf{V}}_n$  still stands for Tyler’s M-estimator. The test  $\phi_{\text{Ty1}}^{(n)}$  has been proposed in Tyler [38]. As shown by the following result, the proposed sign test  $\tilde{\phi}^{(n)}$ , in classical asymptotic scenarios where one stays away from weak identifiability, is asymptotically equivalent to  $\phi_{\text{Ty1}}^{(n)}$  under the null hypothesis.

**Theorem 3.2.** *Let  $(\mathbf{V}_{0n})$  be a sequence of null shape matrices as in (12). Assume that there exists  $\eta > 0$  such that both leading eigenvalues of  $\mathbf{V}_{0n}$  satisfy  $\lambda_{n1}/\lambda_{n2} \geq 1 + \eta$  for  $n$  large enough. Then, under  $\mathbb{P}_{\mathbf{V}_{0n}}^{(n)}$ ,  $T_n(\tilde{\mathbf{V}}_{0n}) = L_n + o_{\mathbb{P}}(1)$  as  $n \rightarrow \infty$ .*

From contiguity, this result also entails that  $\tilde{\phi}^{(n)}$  and  $\phi_{\text{Ty1}}^{(n)}$  are asymptotically equivalent under the local alternatives considered in Theorem 2.1(i), hence exhibit the same asymptotic powers under these. There is no guarantee, however, that this asymptotic equivalence extends to the weak identifiability situations considered in Theorem 2.1(ii)–(iv). To investigate the validity of  $\tilde{\phi}^{(n)}$  under such non-standard asymptotic scenarios, we now thoroughly study the null and non-null asymptotic properties of  $\tilde{\phi}^{(n)}$ .

### 4. Asymptotic properties of the proposed test

We first focus on the null hypothesis. Under the single-spike null hypotheses associated with the shape matrices  $\mathbf{V}_{0n}$  in (2), the test statistic  $T_n(\tilde{\mathbf{V}}_{0n})$  of the feasible test  $\tilde{\phi}^{(n)}$  is asymptotically chi-square with  $p - 1$  degrees of freedom, which easily follows from the asymptotic equivalence result in Theorem 3.1 and from the fact that Theorem 2.1 implies that  $T_n(\mathbf{V}_{0n})$  is asymptotically chi-square with  $p - 1$  degrees of freedom under such sequences. Since the latter theorem focuses on single-spike shape matrices, there is no guarantee, however, that this extends to more general shape matrices. The following result shows that the null asymptotic distribution of  $T_n(\tilde{\mathbf{V}}_{0n})$  remains asymptotically chi-square with  $p - 1$  degrees of freedom for arbitrary sequences of null shape matrices.

**Theorem 4.1.** *Let  $(\mathbf{V}_{0n})$  be a sequence of null shape matrices as in (12). Then, under  $\mathbb{P}_{\mathbf{V}_{0n}}^{(n)}$ ,  $T_n(\tilde{\mathbf{V}}_{0n})$  is asymptotically chi-square with  $p - 1$  degrees of freedom.*

We stress that this result does not only allow for general, multi-spike, null shape matrices, but also for weakly identifiable eigenvectors  $\theta_{n1}$ . In particular, it applies in each of the four asymptotic scenarios considered in Theorem 2.1. Since there is no guarantee that the asymptotic equivalence in Theorem 3.2 holds under weak identifiability, it is unclear whether or not the Tyler test  $\phi_{\text{Tyler}}^{(n)}$  is, like  $\tilde{\phi}^{(n)}$ , robust to weak identifiability. As we will show through simulations in the next section, it actually turns out that the Tyler test severely fails to be robust in this sense.

Now, the nice robustness properties above are not sufficient, on their own, to justify resorting to the proposed sign test, as it might be the case that such robustness is obtained at the expense of power. To see whether or not this is the case, we turn to the investigation of the non-null asymptotic properties of  $\tilde{\phi}^{(n)}$ . We have the following result.

**Theorem 4.2.** Fix  $\theta_0 \in \mathcal{S}^{p-1}$ . Let  $(\delta_n)$  be a positive sequence that is  $O(1)$  and take  $v_n = 1/(\sqrt{n}\gamma_n)$ . Let  $(\tau_n)$  be a sequence converging to  $\tau$  and such that  $\theta_0 + v_n\tau_n \in \mathcal{S}^{p-1}$  for any  $n$ . Then, under  $P_{\mathbf{V}_{1n}}^{(n)}$ , with  $\mathbf{V}_{1n}$  as in (4), we have the following as  $n \rightarrow \infty$ :

- (i) if  $\delta_n \equiv 1$ , then  $T_n(\tilde{\mathbf{V}}_{0n})$  is asymptotically chi-square with  $p - 1$  degrees of freedom and non-centrality parameter

$$\frac{p(p + (p - 1)\xi)}{(p + 2)(p - \xi)} \|\tau\|^2;$$

- (ii) if  $\delta_n$  is  $o(1)$  with  $\sqrt{n}\delta_n \rightarrow \infty$ , then  $T_n(\tilde{\mathbf{V}}_{0n})$  is asymptotically chi-square with  $p - 1$  degrees of freedom and non-centrality parameter

$$\frac{p}{p + 2} \|\tau\|^2;$$

- (iii) if  $\delta_n = 1/\sqrt{n}$ , then  $T_n(\tilde{\mathbf{V}}_{0n})$  is asymptotically chi-square with  $p - 1$  degrees of freedom and non-centrality parameter

$$\frac{p}{p + 2} \|\tau\|^2 \left(1 - \frac{1}{2\xi^2} \|\tau\|^2\right)^2 \left(1 - \frac{1}{4\xi^2} \|\tau\|^2\right);$$

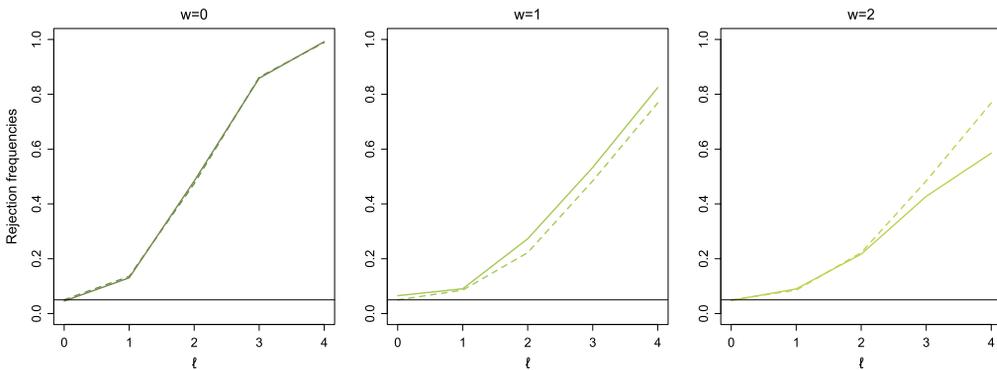
- (iv) if  $\delta_n = 1/\sqrt{n}$ , then, under  $P_{\mathbf{V}_{n1}}^{(n)}$  with  $\mathbf{V}_{n1}$  based on  $\delta_n \equiv 1/\xi$ ,  $T_n(\tilde{\mathbf{V}}_{0n})$  remains asymptotically chi-square with  $p - 1$  degrees of freedom.

From contiguity, Theorem 3.1 implies that the test  $\tilde{\phi}^{(n)}$  enjoys the same asymptotic behavior as  $\phi^{(n)}$  under the contiguous alternatives of Theorem 2.1(i)–(ii), hence inherits the Le Cam optimality properties of the latter test in the corresponding regimes. Consequently, the local asymptotic powers associated with the non-null results in Theorem 4.2(i)–(ii) are the maximal ones that can be achieved. In regime (iii), Theorem 4.2 entails that the test  $\tilde{\phi}^{(n)}$  is rate-optimal, in the sense that it shows non-trivial asymptotic powers against the corresponding contiguous alternatives (the non-standard nature of the limiting experiment in this regime does not allow stating stronger optimality properties, though). Finally, this test is optimal in regime (iv), but trivially so since Theorem 2.1(iv) implies that the trivial  $\alpha$ -test is also optimal in this regime.

We performed the following simulation to check the validity of Theorem 4.2. For any combination of  $\ell \in \{0, 1, 2, 3, 4\}$  and  $w \in \{0, 1, 2\}$ , we generated  $M = 2500$  mutually independent random samples  $\mathbf{X}_1^{(\ell, w)}, \dots, \mathbf{X}_n^{(\ell, w)}$  of size  $n = 200\,000$  from the six-variate ( $p = 6$ ) multinormal distribution with mean zero and covariance matrix

$$\Sigma_n^{(\ell, w)} := \left(1 - \frac{n^{-w/6}}{p}\right) \mathbf{I}_p + n^{-w/6} \boldsymbol{\theta}_{n1}^{(\ell)} \boldsymbol{\theta}_{n1}^{(\ell)'} \tag{17}$$

with  $\boldsymbol{\theta}_0 := (1, 0, \dots, 0)' \in \mathbb{R}^p$  and  $\boldsymbol{\theta}_{n1}^{(\ell)} := (\cos \alpha_{n\ell}, \sin \alpha_{n\ell}, 0, \dots, 0)' \in \mathbb{R}^p$ , where  $\alpha_{n\ell} := 2 \arcsin(\ell v_n/2)$  is based on  $v_n := 1/(\sqrt{n} \gamma_n)$ ; here,  $\gamma_n$  is as in (5), with the values  $\delta_n = n^{-w/6}$  and  $\xi = 1$  that are induced by (17). The value  $w = 0$  yields the standard asymptotic scenario in Theorem 4.2(i), while  $w = 1, 2$  are associated with weak identifiability situations covered by Theorem 4.2(ii). The value  $\ell = 0$  corresponds to the null hypothesis  $\mathcal{H}_0 : \boldsymbol{\theta}_{n1} = \boldsymbol{\theta}_0$ , whereas  $\ell = 1, 2, 3, 4$  provide increasingly severe alternatives of the form  $\boldsymbol{\theta}_{n1}^{(\ell)} = \boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_{n\ell}$ , where  $\boldsymbol{\tau}_{n\ell}$  is some  $p$ -vector with norm  $\ell$ ; this allows us to obtain the corresponding asymptotic local powers from Theorem 4.2(i)–(ii). For each sample, we performed the proposed sign test for  $\mathcal{H}_0 : \boldsymbol{\theta}_{n1} = \boldsymbol{\theta}_0$  in (15) at nominal level 5%. Clearly, the resulting rejection frequencies, that are provided in Figure 1, are in agreement with the theoretical asymptotic powers computed from Theorem 4.2, except maybe for the case  $w = 2$ . Note, however, that at any finite sample size, empirical power curves will eventually converge to flat power curves at the nominal level  $\alpha$  for weak enough identifiability (this follows from Theorem 4.2(iv)), which explains this small deviation observed for  $w = 2$ . The sign nature of the proposed test makes it superfluous to also consider non-Gaussian elliptical distributions here.



**Figure 1.** Rejection frequencies (solid curves), under various null and non-null distributions, of the proposed sign test in (15) performed at asymptotic level  $\alpha = 5\%$ ; the value  $w = 0$  corresponds to the classical asymptotic scenario where  $\boldsymbol{\theta}_{n1}$  remains asymptotically identifiable, whereas  $w = 1, 2$  provide asymptotic scenarios involving weak identifiability (the lighter the color, the weaker the identifiability). Random samples were drawn from six-variate multinormal distributions; see Section 4 for details. Theoretical asymptotic powers are also shown (dashed curves).

### 5. Finite-sample comparisons with competing tests

The objective of this section is to compare the proposed sign test to some competitors. We first focus on empirical size under the null hypothesis. For any  $w \in \{0, 1, 2, 3\}$ , we generated  $M = 5000$  mutually independent random samples  $\mathbf{X}_1^{(w)}, \dots, \mathbf{X}_n^{(w)}$  of size  $n = 400$  from the six-variate ( $p = 6$ ) Gaussian distribution with mean zero and covariance matrix

$$\Sigma_n^{(w)} := \left(1 - \frac{n^{-w/4}}{p}\right) \mathbf{I}_p + n^{-w/4} \boldsymbol{\theta}_0 \boldsymbol{\theta}_0',$$

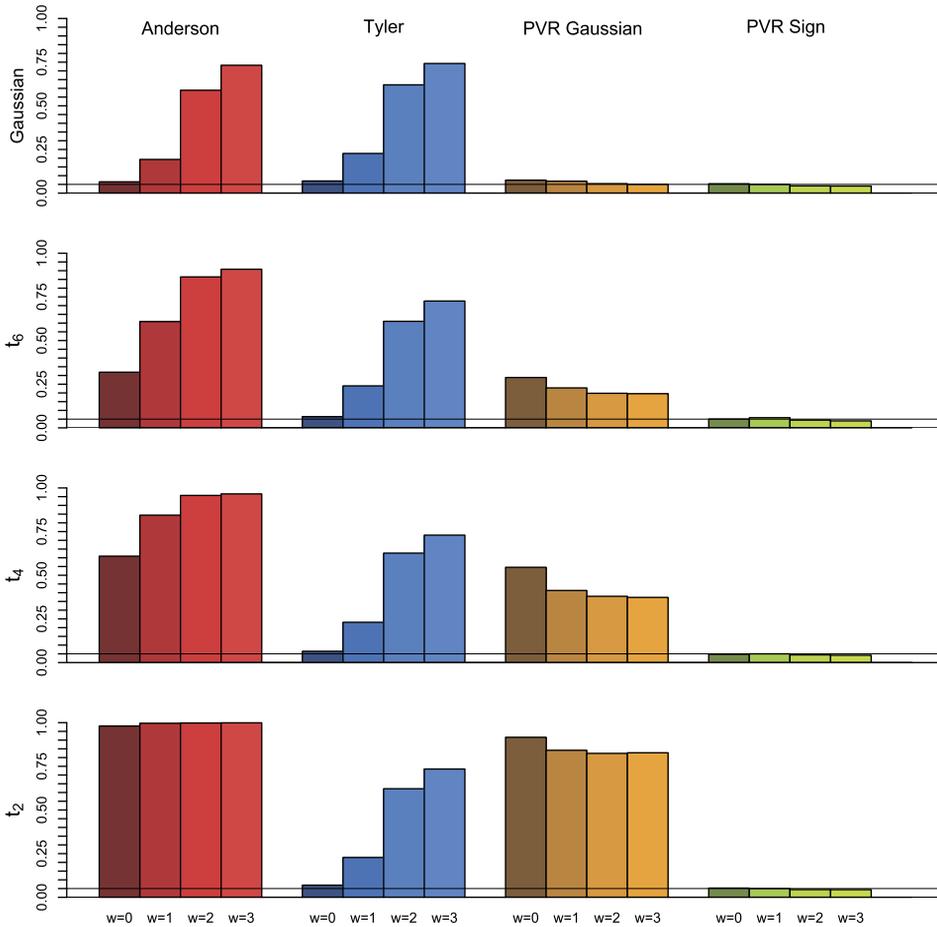
with  $\boldsymbol{\theta}_0 = (1, 0, \dots, 0)' \in \mathbb{R}^p$ . As in the simulation conducted at the end of Section 4, the value  $w = 0$  is associated with classical situations where  $\boldsymbol{\theta}_{n1}$  remains asymptotically identifiable whereas the values  $w = 1, 2, 3$  provide situations where identifiability of  $\boldsymbol{\theta}_{n1}$  is weaker and weaker. In each replication, we performed four tests for  $\mathcal{H}_0 : \boldsymbol{\theta}_{n1} = \boldsymbol{\theta}_0$  at asymptotic level 5%: the classical Gaussian likelihood ratio test from Anderson [1], the Tyler [38] test in (16), the Gaussian test from Paindaveine, Remy and Verdebout [24], and the proposed sign test in (15). The same exercise was repeated with random samples drawn from multivariate  $t$  distributions with 2, 4 and 6 degrees of freedom, in each case with mean zero and scatter matrix  $\Sigma_n^{(w)}$ .

The resulting null rejection frequencies are reported in Figure 2. Clearly, the Anderson [1] test meets the nominal level constraint only in the Gaussian, well identified, case. As expected, the Tyler [38] test, which is a sign test, can deal with heavy tails, but the results make it clear that this test strongly overrejects the null hypothesis under weak identifiability. The opposite holds for the Paindaveine, Remy and Verdebout [24] test, that resists weak identifiability situations in the Gaussian case but cannot deal with heavy tails. In line with the theoretical results of the previous sections, the proposed sign test resists both heavy tails and weak identifiability.

Although the simulation exercise above shows that the proposed sign test is the only one that meets the asymptotic level constraint under heavy tails and weak identifiability, we now turn to a power comparison of the various tests. For  $w \in \{0, 1, 2\}$ ,  $\ell \in \{0, 1, 2, 3\}$  and  $d \in \{1, 2, 3\}$ , we generated  $M = 2500$  mutually independent bivariate ( $p = 2$ ) random samples  $\mathbf{X}_1^{(d,\ell,w)}, \dots, \mathbf{X}_n^{(d,\ell,w)}$  of size  $n = 200$  with covariance/scatter matrix

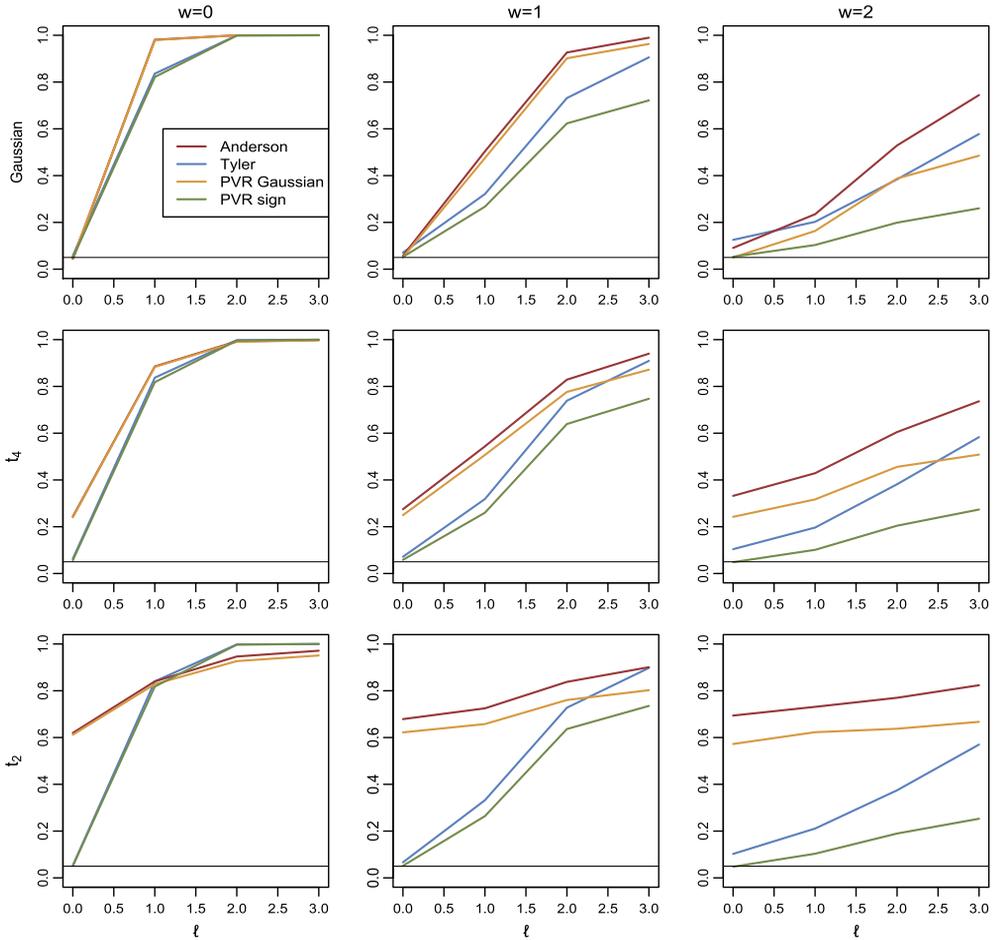
$$\Sigma_n^{(w,\ell)} := \left(1 - \frac{n^{-w/8}}{p}\right) \mathbf{I}_p + n^{-w/8} (\boldsymbol{\theta}_0 + \boldsymbol{\tau}_\ell) (\boldsymbol{\theta}_0 + \boldsymbol{\tau}_\ell)',$$

where  $\boldsymbol{\theta}_0 = (1, 0)'$  and  $\boldsymbol{\tau}_\ell := (\cos(\ell\pi/12) - 1, \sin(\ell\pi/12))'$ . The  $\mathbf{X}_i^{(1,\ell,w)}$ 's have a Gaussian distribution with covariance matrix  $\Sigma_n^{(w,\ell)}$ , the  $\mathbf{X}_i^{(2,\ell,w)}$ 's have a student  $t_4$  distribution with scatter matrix  $\Sigma_n^{(w,\ell)}$ , and the  $\mathbf{X}_i^{(3,\ell,w)}$ 's have a student  $t_2$  distribution with scatter matrix  $\Sigma_n^{(w,\ell)}$ . Note that the value  $\ell = 0$  corresponds to the null hypothesis, while the values  $\ell = 1, 2, 3$  provide increasingly severe alternatives. The parameter  $w$  provides different strengths of identifiability, in the same way as in the previous simulation exercise. We performed the same four tests for  $\mathcal{H}_0 : \boldsymbol{\theta}_{n1} = \boldsymbol{\theta}_0$  as above, still at asymptotic level 5%. The resulting rejection frequencies are



**Figure 2.** Null rejection frequencies, under six-variate Gaussian,  $t_6$ ,  $t_4$  and  $t_2$  densities, of four tests for the null hypothesis  $\mathcal{H}_0 : \theta_{n1} = (1, 0, \dots, 0)' (\in \mathbb{R}^6)$ , all performed at asymptotic level  $\alpha = 5\%$ . The tests considered are the Anderson [1] test, the Tyler [38] test, the Gaussian test from Paindaveine, Remy and Verdebout [24], and the proposed sign test;  $w = 0$  corresponds to the classical asymptotic scenario where  $\theta_{n1}$  remains asymptotically identifiable, whereas  $w = 1, 2, 3$  provide asymptotic scenarios involving weaker and weaker identifiability; see Section 5 for details.

plotted in Figure 3. Clearly, the results confirm that the proposed sign test is the only test that is robust to both weak identifiability and tail heaviness. It is also seen that this sign test shows power under weakly identified situations and that this power, as expected, does not depend on the tails of the parent distribution. Finally, note that the weaker the identifiability, the larger the sample size needs to be to provide some power, which is quite natural.



**Figure 3.** Empirical power curves, under bivariate Gaussian (top),  $t_4$  densities (middle) and  $t_2$  densities (bottom), of four tests for the null hypothesis  $\mathcal{H}_0 : \theta_{n1} = (1, 0)'$ , all performed at asymptotic level  $\alpha = 5\%$ . The tests considered are the Anderson [1] test, the Tyler [38] test, the Gaussian test from Paindaveine, Remy and Verdebout [24], and the proposed sign test;  $w = 0$  corresponds to the classical asymptotic scenario where  $\theta_{n1}$  remains asymptotically identifiable, whereas  $w = 1$  and  $w = 2$  provide asymptotic scenarios involving weaker identifiability; see Section 5 for details.

## 6. Real data illustration

We illustrate the practical relevance of the proposed sign test on the famous Swiss banknote dataset, that was also used for illustration in Paindaveine, Remy and Verdebout [24], to which we refer for more details. This dataset is available in the R package `uskewfactors` (Murray, Browne and McNicholas [20]) and consists of six measurements on 100 genuine and 100 counterfeit old Swiss 1000-franc banknotes. As in Flury [8] (see pages 41–43), we restrict here to

$n = 85$  counterfeit bills made by the same forger and focus on four of the six available measurements: the width  $L$  of the left side of the banknote, the width  $R$  on its right side, the width  $B$  of the bottom margin and the width  $T$  of the top margin, all measured in  $\text{mm} \times 10^{-1}$ . The resulting sample covariance matrix has eigenvalues  $\hat{\lambda}_1 = 101.48$ ,  $\hat{\lambda}_2 = 12.89$ ,  $\hat{\lambda}_3 = 10.11$  and  $\hat{\lambda}_4 = 2.63$ , and corresponding eigenvectors

$$\hat{\theta}_1 = \begin{pmatrix} 0.032 \\ -0.012 \\ 0.820 \\ -0.571 \end{pmatrix}, \quad \hat{\theta}_2 = \begin{pmatrix} 0.593 \\ 0.797 \\ 0.057 \\ 0.097 \end{pmatrix}, \quad \hat{\theta}_3 = \begin{pmatrix} -0.015 \\ -0.129 \\ 0.566 \\ 0.814 \end{pmatrix}, \quad \text{and} \quad \hat{\theta}_4 = \begin{pmatrix} 0.804 \\ -0.590 \\ -0.064 \\ -0.035 \end{pmatrix}.$$

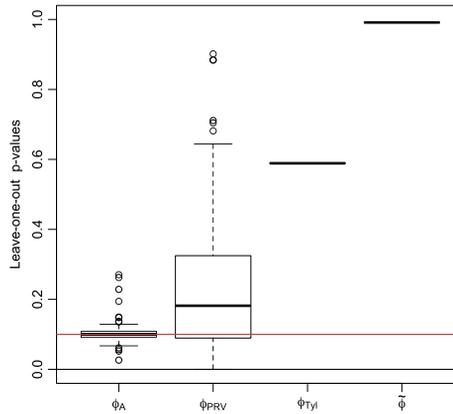
Clearly, the first principal component can be interpreted as the vertical position of the print image on the bill since it is a contrast between  $B$  and  $T$ . Similarly, the second principal component could be interpreted as an aggregate of  $L$  and  $R$ , that is, as the vertical size of the bill. Yet, Flury refrains from interpreting the second component in this way as the second and third roots are quite close to each other. Accordingly, he reports that the corresponding eigenvectors should be considered spherical.

In view of the discussion above, it is natural to test that  $L$  and  $R$  indeed contribute equally to the second component and that no other variables contribute to it. In other words, it is natural to test the null hypothesis  $\mathcal{H}_0 : \theta_2 = \theta_2^0$ , with  $\theta_2^0 := (1, 1, 0, 0)' / \sqrt{2}$ . While the tests discussed in the present paper address testing problems on the first eigenvector  $\theta_1$ , obvious modifications of these tests allow performing inference on any other eigenvector  $\theta_j$ ,  $j = 2, \dots, p$ . In Paindaveine, Remy and Verdebout [24], the Anderson test  $\phi_A$  and the Gaussian test  $\phi_{\text{PVR}}$  proposed in that paper were used to test the null hypothesis  $\mathcal{H}_0 : \theta_2 = \theta_2^0$ . It is well-known, however, that several observations in the dataset may be considered as outliers (see, e.g., Salibián-Barrera, Van Aelst and Willems [29]), which motivates us resorting to robust tests such as the Tyler test  $\phi_{\text{TYI}}$  and our sign test  $\tilde{\phi}$ . When testing the null hypothesis  $\mathcal{H}_0$  above, these robust tests provided  $p$ -values 0.609 and 0.992, respectively, which is to be compared to the  $p$ -values 0.099 and 0.177, respectively provided by  $\phi_A$  and  $\phi_{\text{PVR}}$ . This shows that, at level 10%, only the Anderson test leads to rejection of the null hypothesis. The robustness of our sign test and the fact that the Anderson test tends to strongly overreject the null hypothesis under weak identifiability should make practitioners confident that non-rejection is the right decision in the present case.

To complement the analysis, we performed the same four tests on the 85 subsamples of size 84 obtained by removing one observation from the sample considered above. Figure 4 provides, for each test, a boxplot of the resulting 85 “leave-one-out”  $p$ -values. The results show that the Anderson test rejects the null hypothesis much more often than the other tests. It is remarkable that the tests based on spatial signs never led to rejection at any usual nominal level, which, arguably, is due to the natural robustness of spatial signs and of the Tyler estimator of shape.

## 7. Conclusions and final comments

In this work, we considered hypothesis testing for principal directions in challenging asymptotic scenarios involving weak identifiability. Under ellipticity assumptions, we proposed a sign test



**Figure 4.** Boxplots of the 85 “leave-one-out”  $p$ -values of the Anderson test ( $\phi_A$ ), of the Gaussian Paindaveine, Remy and Verdebout [24] test ( $\phi_{\text{PVR}}$ ), of the Tyler [38] test ( $\phi_{\text{Tyl}}$ ), and of the proposed sign test ( $\tilde{\phi}$ ), when testing the null hypothesis  $\mathcal{H}_0 : \theta_2 := (1, 1, 0, 0)' / \sqrt{2}$ . More precisely, these  $p$ -values are those obtained when applying the corresponding tests to the 85 subsample of size 84 obtained by removing one observation in the real data set considered in the PCA analysis of Flury [8], pages 41–43.

that, unlike its competitors, meets the asymptotic level constraint both under heavy tails and under weak identifiability. By resorting to Le Cam’s asymptotic theory of statistical experiments, we also proved that this test enjoys strong optimality properties in the class of spatial sign tests (all optimality statements below are relative to this class of tests). In particular, it is locally asymptotically optimal in classical situations where  $\theta_{n1}$  remains asymptotically identifiable. It follows from our results that the likelihood ratio test from Tyler [38] satisfies the same optimality property. Our sign test, however, shows strong advantages over the Tyler test: not only does our test meet the level constraint under *any* weak identifiability situation, but it also remains locally asymptotically optimal in all cases, but for the case  $\delta_n \sim 1/\sqrt{n}$  for which our test is still rate-optimal.

The following comments are in order. First, our sign test is not only robust to heavy tails and weak identifiability but also to (some) departures from ellipticity. More precisely, it should be clear that our sign test only assumes that the spatial signs  $\mathbf{U}_{ni} = \mathbf{X}_{ni} / \|\mathbf{X}_{ni}\|$  follow an angular Gaussian distribution. As a consequence, the spatial signs do not need be independent of the radii  $\|\mathbf{X}_{ni}\|$ , which implies in particular that our sign test can deal with some skewed distributions. More precisely, the proposed test only requires that the observations  $\mathbf{X}_{ni}$ ,  $i = 1, \dots, n$ , form a random sample from a distribution *with elliptical directions*; see Randles [28]. Second, it has been throughout assumed that the parent elliptical distribution was centered. This was mainly for the sake of readability, as our results can easily be extended to the unspecified location case. More precisely, an unspecified-location version of the proposed test can simply be obtained by replacing the spatial signs  $\mathbf{U}_{ni} = \mathbf{X}_{ni} / \|\mathbf{X}_{ni}\|$  in our sign test with centered versions  $\mathbf{U}_{ni}(\hat{\boldsymbol{\mu}}_n) = (\mathbf{X}_{ni} - \hat{\boldsymbol{\mu}}_n) / \|\mathbf{X}_{ni} - \hat{\boldsymbol{\mu}}_n\|$ , where  $\hat{\boldsymbol{\mu}}_n$  is an arbitrary root- $n$  consistent estimator of the center of the underlying elliptical distribution. A natural choice, that would be root- $n$  consistent even in the large class of distributions with elliptical directions, is the affine-equivariant median from

Hettmansperger and Randles [14]. Due to the (Fisher) orthogonality between location and scatter parameters under ellipticity (see Hallin and Paindaveine [11]), all asymptotic results of this paper readily extend to the resulting unspecified-location sign test.

While this work provides an overall good procedure to test for principal directions under weak identifiability, it also opens perspectives for future research. As mentioned above, the optimality of the proposed test is relative to the class of spatial sign tests. Restricting to sign tests of course is a guarantee for excellent robustness properties, yet it might be so that, if some slightly lower robustness is also acceptable, then higher asymptotic efficiency could be achieved. In particular, it should be possible to develop signed rank tests that provide a nice trade-off between efficiency and robustness. It is expected that such tests can deal with heavy tails and are robust to weak identifiability, while uniformly dominating, in terms of asymptotic relative efficiencies, parametric Gaussian tests in classical cases where the leading principal direction remains asymptotically identifiable; see Paindaveine [22].

## Appendix A: Proof of Theorem 2.1

We start with some preliminary lemmas.

**Lemma A.1.** *Consider the shape matrices  $\mathbf{V}_{0n}$  and  $\mathbf{V}_{1n}$  in (2) and (4). Then, (i)  $\mathbf{V}_{0n}$  and  $\mathbf{V}_{1n}$  share the same determinant; (ii) for any real number  $a$ , the  $a$ th matrix powers of  $\mathbf{V}_{0n}$  and  $\mathbf{V}_{1n}$  are given by*

$$\begin{aligned}\mathbf{V}_{0n}^a &= \left(1 - \frac{\delta_n \xi}{p}\right)^a (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) + \left(1 + \frac{(p-1)\delta_n \xi}{p}\right)^a \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0 \\ &= \left(1 - \frac{\delta_n \xi}{p}\right)^a \mathbf{I}_p + \lambda_{a,n} \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0\end{aligned}$$

and

$$\begin{aligned}\mathbf{V}_{1n}^a &= \left(1 - \frac{\delta_n \xi}{p}\right)^a (\mathbf{I}_p - (\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n)(\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n)') + \lambda_{a,n} (\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n)(\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n)' \\ &= \left(1 - \frac{\delta_n \xi}{p}\right)^a \mathbf{I}_p + \lambda_{a,n} (\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n)(\boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n)',\end{aligned}$$

where we let  $\lambda_{a,n} := (1 + (p-1)\delta_n \xi/p)^a - (1 - \delta_n \xi/p)^a$ .

**Proof of Lemma A.1.** (i) Letting  $\boldsymbol{\theta}_n := \boldsymbol{\theta}_0 + \nu_n \boldsymbol{\tau}_n$ ,  $r := 1 - \delta_n \xi/p$  and  $s := 1 + (p-1)\delta_n \xi/p$ , rewrite  $\mathbf{V}_{0n}$  and  $\mathbf{V}_{1n}$  as

$$\mathbf{V}_{0n} = r(\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) + s \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0 \quad \text{and} \quad \mathbf{V}_{1n} = r(\mathbf{I}_p - \boldsymbol{\theta}_n \boldsymbol{\theta}'_n) + s \boldsymbol{\theta}_n \boldsymbol{\theta}'_n. \quad (18)$$

Both matrices have eigenvalues  $r$  with multiplicity  $p-1$  and  $s$  with multiplicity one, hence have determinant  $r^{p-1}s$ . (ii) The result directly follows from the spectral decompositions in (18).  $\square$

**Lemma A.2.** *If  $\delta_n \equiv 1$ , then  $\lim_{n \rightarrow \infty} p\delta_n\xi/\gamma_n = p + (p - 1)\xi$ , where  $\gamma_n$  is defined in (5). If  $\delta_n = o(1)$ , then  $\lim_{n \rightarrow \infty} p\delta_n\xi/\gamma_n = p$ .*

**Proof of Lemma A.2.** Since  $p\delta_n\xi/\gamma_n = p + (p - 1)\delta_n\xi$ , the result follows. □

To state the next result, let

$$\mathbf{T}_n(\mathbf{V}) := \frac{1}{n} \sum_{i=1}^n \text{vec}(\mathbf{U}_{ni}(\mathbf{V})\mathbf{U}'_{ni}(\mathbf{V}))(\text{vec}(\mathbf{U}_{ni}(\mathbf{V})\mathbf{U}'_{ni}(\mathbf{V})))',$$

with  $\mathbf{U}_{ni}(\mathbf{V}) := \mathbf{V}^{-1/2}\mathbf{U}_{ni}/\|\mathbf{V}^{-1/2}\mathbf{U}_{ni}\|$ ; note that with this notation,  $\mathbf{S}_n(\mathbf{V})$  in (5) rewrites

$$\mathbf{S}_n(\mathbf{V}) := \frac{1}{n} \sum_{i=1}^n \mathbf{U}_{ni}(\mathbf{V})\mathbf{U}'_{ni}(\mathbf{V}).$$

We then have the following result.

**Lemma A.3.** *Fix an arbitrary sequence of shape matrices  $(\mathbf{V}_n)$ . Then, under  $P_{\mathbf{V}_n}^{(n)}$ ,*

$$(i) \quad \mathbf{T}_n(\mathbf{V}_n) = \frac{1}{p(p+2)}(\mathbf{I}_p + \mathbf{K}_p + \mathbf{J}_p) + o_P(1)$$

and

$$(ii) \quad \sqrt{n} \text{vec}\left(\mathbf{S}_n(\mathbf{V}_n) - \frac{1}{p}\mathbf{I}_p\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \frac{1}{p(p+2)}(\mathbf{I}_p + \mathbf{K}_p + \mathbf{J}_p) - \frac{1}{p^2}\mathbf{J}_p\right)$$

as  $n \rightarrow \infty$ .

**Proof of Lemma A.3.** For any  $n$ ,  $\mathbf{U}_{n1}(\mathbf{V}_n), \dots, \mathbf{U}_{nn}(\mathbf{V}_n)$  form a random sample from the uniform distribution on  $S^{p-1}$ . Therefore, the result follows from (i) the weak law of large numbers and from (ii) the central limit theorem, by using in both cases Lemma A.2 in Paindaveine and Verdebout [25]. □

**Proof of Theorem 2.1.** First note that the quantity  $\gamma_n$  in (5) satisfies

$$-\gamma_n = \left(1 - \frac{\delta_n\xi}{p}\right)\lambda_{-1,n}, \tag{19}$$

where  $\lambda_{-1,n}$  was defined in Lemma A.1. Part (ii) of this lemma therefore yields

$$\mathbf{U}'_{ni}\mathbf{V}_{0n}^{-1}\mathbf{U}_{ni} = \left(1 - \frac{\delta_n\xi}{p}\right)^{-1} + \lambda_{-1,n}(\mathbf{U}'_{ni}\boldsymbol{\theta}_0)^2 = \left(1 - \frac{\delta_n\xi}{p}\right)^{-1} (1 - \gamma_n(\mathbf{U}'_{ni}\boldsymbol{\theta}_0)^2) \tag{20}$$

and, similarly,

$$\mathbf{U}'_{ni}\mathbf{V}_{n1}^{-1}\mathbf{U}_{ni} = \left(1 - \frac{\delta_n\xi}{p}\right)^{-1} (1 - \gamma_n(\mathbf{U}'_{ni}(\boldsymbol{\theta}_0 + v_n\boldsymbol{\tau}_n))^2).$$

Recalling the angular density in (1) and using Lemma A.1(i), we then obtain

$$\begin{aligned} \Lambda_n &= -\frac{p}{2} \sum_{i=1}^n \{ \log(\mathbf{U}'_{ni} \mathbf{V}_{1n}^{-1} \mathbf{U}_{ni}) - \log(\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}) \} \\ &= -\frac{p}{2} \sum_{i=1}^n \{ \log(1 - \gamma_n(\mathbf{U}'_{ni}(\boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_n))^2) - \log(1 - \gamma_n(\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2) \}, \end{aligned}$$

which, by writing  $\gamma_n(\mathbf{U}'_{ni}(\boldsymbol{\theta}_0 + v_n \boldsymbol{\tau}_n))^2 = \gamma_n(\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2 + 2\gamma_n v_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)(\mathbf{U}'_{ni} \boldsymbol{\tau}_n) + \gamma_n v_n^2 (\mathbf{U}'_{ni} \boldsymbol{\tau}_n)^2$ , yields

$$\begin{aligned} \Lambda_n &= -\frac{p}{2} \sum_{i=1}^n \log \left( 1 - \frac{2\gamma_n v_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)(\mathbf{U}'_{ni} \boldsymbol{\tau}_n) + \gamma_n v_n^2 (\mathbf{U}'_{ni} \boldsymbol{\tau}_n)^2}{1 - \gamma_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2} \right) \\ &= -\frac{p}{2} \sum_{i=1}^n \log \left( 1 - \gamma_n v_n \frac{2(\mathbf{U}'_{ni} \boldsymbol{\theta}_0)(\mathbf{U}'_{ni} \boldsymbol{\tau}_n) + v_n (\mathbf{U}'_{ni} \boldsymbol{\tau}_n)^2}{1 - \gamma_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2} \right) \\ &=: -\frac{p}{2} \sum_{i=1}^n \log(1 - \gamma_n v_n R_{ni}). \end{aligned} \tag{21}$$

A Taylor expansion yields

$$\begin{aligned} \Lambda_n &= \left( \frac{p\gamma_n v_n}{2} \sum_{i=1}^n R_{ni} \right) + \left( \frac{p\gamma_n^2 v_n^2}{4} \sum_{i=1}^n R_{ni}^2 \right) + \left( \frac{p\gamma_n^3 v_n^3}{6} \sum_{i=1}^n \frac{R_{ni}^3}{(1 - \gamma_n v_n H_{ni})^3} \right) \\ &=: L_{n1} + L_{n2} + L_{n3}, \end{aligned}$$

for some  $H_{ni}$  between 0 and  $R_{ni}$ . Note that, in all cases (i)–(iv) considered in the theorem, we have that  $v_n \gamma_n = O(1/\sqrt{n})$ ,  $v_n = O(1)$  and that there exists  $\eta \in (0, 1)$  such that  $|\gamma_n| < 1 - \eta$  (recall that, in case (i),  $\xi$  is chosen in such a way that  $\mathbf{V}_{0n}$  is positive definite). Consequently, the boundedness of the sequence  $(\boldsymbol{\tau}_n)$ , along with the fact that  $\|\mathbf{U}_{ni}\| = 1$  almost surely, ensures that there exists a positive constant  $C$  such that  $|L_{n3}| \leq C/\sqrt{n}$  almost surely, so that  $L_{n3}$  is  $op(1)$  (all stochastic convergences in this proof are as  $n \rightarrow \infty$  under  $\mathbb{P}_{\mathbf{V}_{0n}}^{(n)}$ ).

Now, using (20), rewrite  $L_{n1}$  as

$$\begin{aligned} L_{n1} &= \frac{p\gamma_n v_n}{2} \sum_{i=1}^n \frac{2(\mathbf{U}'_{ni} \boldsymbol{\theta}_0)(\mathbf{U}'_{ni} \boldsymbol{\tau}_n) + v_n (\mathbf{U}'_{ni} \boldsymbol{\tau}_n)^2}{1 - \gamma_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2} \\ &= p\gamma_n v_n \sum_{i=1}^n \frac{\boldsymbol{\tau}'_n \mathbf{U}_{ni} \mathbf{U}'_{ni} (\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n)}{1 - \gamma_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2} \end{aligned}$$

$$\begin{aligned}
 &= p\gamma_n v_n \left(1 - \frac{\delta_n \xi}{p}\right)^{-1} \sum_{i=1}^n \frac{\boldsymbol{\tau}'_n \mathbf{U}_{ni} \mathbf{U}'_{ni} (\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n)}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \\
 &= \frac{p^2 n \gamma_n v_n}{p - \delta_n \xi} \boldsymbol{\tau}'_n \mathbf{V}_{0n}^{1/2} \mathbf{S}_n (\mathbf{V}_{0n}) \mathbf{V}_{0n}^{1/2} \left(\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n\right).
 \end{aligned}$$

Note that, using (3), we have

$$\begin{aligned}
 \boldsymbol{\tau}'_n \mathbf{V}_{0n} \left(\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n\right) &= \left(1 - \frac{\delta_n \xi}{p}\right) \boldsymbol{\tau}'_n \left(\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n\right) + \delta_n \xi (\boldsymbol{\tau}'_n \boldsymbol{\theta}_0) \boldsymbol{\theta}'_0 \left(\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n\right) \\
 &= \delta_n \xi (\boldsymbol{\tau}'_n \boldsymbol{\theta}_0) \left(1 + \frac{1}{2} v_n \boldsymbol{\theta}'_0 \boldsymbol{\tau}_n\right) \\
 &= -\frac{1}{2} v_n \delta_n \xi \|\boldsymbol{\tau}_n\|^2 \left(1 + \frac{1}{2} v_n \boldsymbol{\theta}'_0 \boldsymbol{\tau}_n\right),
 \end{aligned}$$

which yields

$$\begin{aligned}
 L_{n1} &= \frac{p^2 n \gamma_n v_n}{p - \delta_n \xi} \boldsymbol{\tau}'_n \mathbf{V}_{0n}^{1/2} \left(\mathbf{S}_n (\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p\right) \mathbf{V}_{0n}^{1/2} \left(\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n\right) \\
 &\quad + \frac{pn \gamma_n v_n}{p - \delta_n \xi} \boldsymbol{\tau}'_n \mathbf{V}_{0n} \left(\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n\right) \\
 &= \frac{p^2 n \gamma_n v_n}{p - \delta_n \xi} \boldsymbol{\tau}'_n \mathbf{V}_{0n}^{1/2} \left(\mathbf{S}_n (\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p\right) \mathbf{V}_{0n}^{1/2} \left(\boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n\right) \\
 &\quad - \frac{pn \gamma_n v_n^2 \delta_n \xi}{2(p - \delta_n \xi)} \|\boldsymbol{\tau}_n\|^2 \left(1 + \frac{1}{2} v_n \boldsymbol{\theta}'_0 \boldsymbol{\tau}_n\right), \tag{22}
 \end{aligned}$$

where, from Lemma A.1, we have

$$\begin{aligned}
 \mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0 &= \left(1 + \frac{(p-1)\delta_n \xi}{p}\right)^{1/2} \boldsymbol{\theta}_0 \quad \text{and} \\
 \mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n &= \left(1 - \frac{\delta_n \xi}{p}\right)^{1/2} \boldsymbol{\tau}_n + \lambda_{-1/2,n} (\boldsymbol{\theta}'_0 \boldsymbol{\tau}_n) \boldsymbol{\theta}_0.
 \end{aligned} \tag{23}$$

Turning to  $L_{n2}$ ,

$$\begin{aligned}
 L_{n2} &= \frac{p\gamma_n^2 v_n^2}{4} \sum_{i=1}^n \frac{\{2(\mathbf{U}'_{ni} \boldsymbol{\theta}_0)(\mathbf{U}'_{ni} \boldsymbol{\tau}_n) + v_n (\mathbf{U}'_{ni} \boldsymbol{\tau}_n)^2\}^2}{(1 - \gamma_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2)^2} \\
 &= \frac{p\gamma_n^2 v_n^2}{4} \left\{4(\boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0)' \sum_{i=1}^n \frac{\text{vec}(\mathbf{U}_{ni} \mathbf{U}'_{ni})(\text{vec}(\mathbf{U}_{ni} \mathbf{U}'_{ni}))'}{(1 - \gamma_n (\mathbf{U}'_{ni} \boldsymbol{\theta}_0)^2)^2} (\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n)\right\}
 \end{aligned}$$

$$\begin{aligned}
 &+ 4v_n(\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n)' \sum_{i=1}^n \frac{\text{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni})(\text{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni}))'}{(1 - \gamma_n(\mathbf{U}'_{ni}\boldsymbol{\theta}_0)^2)^2} (\boldsymbol{\tau}_n \otimes \boldsymbol{\theta}_0) \\
 &+ v_n^2(\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n)' \sum_{i=1}^n \frac{\text{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni})(\text{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni}))'}{(1 - \gamma_n(\mathbf{U}'_{ni}\boldsymbol{\theta}_0)^2)^2} (\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n) \Big\}.
 \end{aligned}$$

Using (20) again, Lemma A.3(i), and the fact that  $(\mathbf{v} \otimes \mathbf{v})' \mathbf{K}_p = \mathbf{K}_1(\mathbf{v} \otimes \mathbf{v})' = (\mathbf{v} \otimes \mathbf{v})'$  for any  $p$ -vector  $\mathbf{v}$ , this yields

$$\begin{aligned}
 L_{n2} &= \frac{np\gamma_n^2 v_n^2}{4} \left(1 - \frac{\delta_n \xi}{p}\right)^{-2} \left\{ 4((\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0))' \mathbf{T}_n(\mathbf{V}_{0n})((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n)) \right. \\
 &\quad + 4v_n((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n))' \mathbf{T}_n(\mathbf{V}_{0n})((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0)) \\
 &\quad \left. + v_n^2((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n))' \mathbf{T}_n(\mathbf{V}_{0n})((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n)) \right\} \\
 &= \frac{np^2 \gamma_n^2 v_n^2}{4(p+2)(p - \delta_n \xi)^2} \left\{ 4((\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0))' (2\mathbf{I}_p + \mathbf{J}_p)((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n)) \right. \\
 &\quad + 4v_n((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n))' (2\mathbf{I}_p + \mathbf{J}_p)((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0)) \\
 &\quad \left. + v_n^2((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n))' (2\mathbf{I}_p + \mathbf{J}_p)((\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n)) + o_P(1) \right\}. \tag{24}
 \end{aligned}$$

We can now consider the cases (i)–(iv). We start with cases (i)–(ii) and let  $\zeta$  be equal to one if case (i) is considered and to zero if case (ii) is. Since  $v_n \gamma_n = 1/\sqrt{n}$  in both cases, we have

$$\begin{aligned}
 L_{n1} &= \frac{p^2 \sqrt{n}}{p - \delta_n \xi} \boldsymbol{\tau}'_n \mathbf{V}_{0n}^{1/2} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \mathbf{V}_{0n}^{1/2} \left( \boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n \right) \\
 &\quad - \frac{p \delta_n \xi}{2\gamma_n(p - \delta_n \xi)} \|\boldsymbol{\tau}_n\|^2 \left( 1 + \frac{1}{2} v_n \boldsymbol{\theta}'_0 \boldsymbol{\tau}_n \right) \\
 &= \frac{p^2 \sqrt{n}}{p - \zeta \xi} \boldsymbol{\tau}'_n \mathbf{V}_{0n}^{1/2} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \mathbf{V}_{0n}^{1/2} \left( \boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n \right) \\
 &\quad - \frac{p + \zeta(p - 1)\xi}{2(p - \zeta \xi)} \|\boldsymbol{\tau}_n\|^2 + o_P(1),
 \end{aligned}$$

where the last equality follows from Lemma A.3(ii), Lemma A.2, and the fact that  $v_n = o(1)$ .

So, using (23), Lemma A.3 again, and the fact that  $\boldsymbol{\theta}'_0 \boldsymbol{\tau}_n = o(1)$ , we obtain

$$\begin{aligned}
 L_{n1} &= \frac{p^2 \sqrt{n}}{p - \zeta \xi} \left(1 - \frac{\zeta \xi}{p}\right)^{1/2} \left(1 + \frac{\zeta(p - 1)\xi}{p}\right)^{1/2} \boldsymbol{\tau}'_n \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \boldsymbol{\theta}_0 \\
 &\quad - \frac{p + \zeta(p - 1)\xi}{2(p - \zeta \xi)} \|\boldsymbol{\tau}_n\|^2 + o_P(1)
 \end{aligned}$$

$$\begin{aligned}
&= \frac{p(p + \zeta(p - 1)\xi)^{1/2}\sqrt{n}}{(p - \zeta\xi)^{1/2}} \boldsymbol{\tau}'_n \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \boldsymbol{\theta}_0 - \frac{p + \zeta(p - 1)\xi}{2(p - \zeta\xi)} \|\boldsymbol{\tau}_n\|^2 + o_{\mathbb{P}}(1) \\
&= \frac{p(p + \zeta(p - 1)\xi)^{1/2}\sqrt{n}}{(p - \zeta\xi)^{1/2}} \boldsymbol{\tau}'_n (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \boldsymbol{\theta}_0 \\
&\quad - \frac{p + \zeta(p - 1)\xi}{2(p - \zeta\xi)} \boldsymbol{\tau}'_n (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \boldsymbol{\tau}_n + o_{\mathbb{P}}(1). \tag{25}
\end{aligned}$$

Now, in cases (i)–(ii), since  $v_n = 1/(\sqrt{n}\gamma_n) = o(1)$ , (24) becomes

$$L_{n2} = \frac{p^2}{(p + 2)(p - \zeta\xi)^2} \left( (\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\theta}_0) \right)' (2\mathbf{I}_p + \mathbf{J}_p) \left( (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \otimes (\mathbf{V}_{0n}^{1/2} \boldsymbol{\tau}_n) \right) + o_{\mathbb{P}}(1),$$

which, by using (23), readily yields

$$\begin{aligned}
L_{n2} &= \frac{p^2}{(p + 2)(p - \zeta\xi)^2} \left( 1 + \frac{(p - 1)\delta_n \xi}{p} \right) \left( 1 - \frac{\delta_n \xi}{p} \right) (\boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0)' (2\mathbf{I}_p + \mathbf{J}_p) (\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n) \\
&\quad + o_{\mathbb{P}}(1) \\
&= \frac{p + \zeta(p - 1)\xi}{(p + 2)(p - \zeta\xi)} \{ 2(\boldsymbol{\theta}'_0 \boldsymbol{\tau}_n)^2 + \|\boldsymbol{\tau}_n\|^2 \} + o_{\mathbb{P}}(1) \\
&= \frac{p + \zeta(p - 1)\xi}{(p + 2)(p - \zeta\xi)} \boldsymbol{\tau}'_n (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \boldsymbol{\tau}_n + o_{\mathbb{P}}(1). \tag{26}
\end{aligned}$$

The results in (6) and (7) then follow from (25)–(26), whereas the asymptotic normality results for  $\boldsymbol{\Delta}_{\boldsymbol{\theta}_0}^{(i)}$  and  $\boldsymbol{\Delta}_{\boldsymbol{\theta}_0}^{(ii)}$  are direct corollaries of Lemma A.3(ii).

We turn to case (iii), for which  $\delta_n = 1/\sqrt{n}$  and  $v_n = 1/(\sqrt{n}\gamma_n) = 1/\xi + o(1)$ . By using Lemma A.2, (22) here becomes

$$\begin{aligned}
L_{n1} &= \frac{p}{p - (\xi/\sqrt{n})} \boldsymbol{\tau}'_n \mathbf{V}_{0n}^{1/2} \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \mathbf{V}_{0n}^{1/2} \left( \boldsymbol{\theta}_0 + \frac{1}{2} v_n \boldsymbol{\tau}_n \right) \\
&\quad - \frac{p\delta_n \xi}{2\gamma_n(p - (\xi/\sqrt{n}))} \|\boldsymbol{\tau}_n\|^2 \left( 1 + \frac{1}{2} v_n \boldsymbol{\theta}'_0 \boldsymbol{\tau}_n \right) \\
&= \boldsymbol{\tau}'_n \mathbf{V}_{0n}^{1/2} \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \mathbf{V}_{0n}^{1/2} \left( \boldsymbol{\theta}_0 + \frac{1}{2\xi} \boldsymbol{\tau}_n \right) - \frac{1}{2} \|\boldsymbol{\tau}_n\|^2 \left( 1 - \frac{1}{4\xi^2} \|\boldsymbol{\tau}_n\|^2 \right) + o_{\mathbb{P}}(1).
\end{aligned}$$

Since  $\mathbf{V}_{0n} = \mathbf{I}_p + o(1)$ , this yields

$$L_{n1} = \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\theta}_0 + \frac{1}{2\xi} \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\tau}_n - \frac{1}{2} \|\boldsymbol{\tau}_n\|^2 + \frac{1}{8\xi^2} \|\boldsymbol{\tau}_n\|^4 + o_{\mathbb{P}}(1). \tag{27}$$

Turning to  $L_{n2}$ , (24) provides

$$\begin{aligned}
 L_{n2} &= \frac{p^2}{4(p+2)(p+o(1))^2} \{4(\boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0)'(2\mathbf{I}_p + \mathbf{J}_p)(\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n) \\
 &\quad + 4\xi^{-1}(\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n)'(2\mathbf{I}_p + \mathbf{J}_p)(\boldsymbol{\tau}_n \otimes \boldsymbol{\theta}_0) + \xi^{-2}(\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n)'(2\mathbf{I}_p + \mathbf{J}_p)(\boldsymbol{\tau}_n \otimes \boldsymbol{\tau}_n)\} + o_P(1) \\
 &= \frac{1}{4(p+2)} \{8(\boldsymbol{\theta}'_0 \boldsymbol{\tau}_n)^2 + 4\|\boldsymbol{\tau}_n\|^2 + 12\xi^{-1}(\boldsymbol{\theta}'_0 \boldsymbol{\tau}_n)\|\boldsymbol{\tau}_n\|^2 + 3\xi^{-2}\|\boldsymbol{\tau}_n\|^4\} + o_P(1) \\
 &= \frac{1}{p+2} \|\boldsymbol{\tau}_n\|^2 - \frac{1}{4(p+2)\xi^2} \|\boldsymbol{\tau}_n\|^4 + o_P(1). \tag{28}
 \end{aligned}$$

Hence, from (27)–(28), we conclude that

$$\Lambda_n = \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\theta}_0 + \frac{1}{2\xi} \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\tau}_n - \frac{P}{2(p+2)} \|\boldsymbol{\tau}_n\|^2 + \frac{P}{8(p+2)\xi^2} \|\boldsymbol{\tau}_n\|^4 + o_P(1),$$

as was to be shown. Again, the asymptotic normality result for  $\text{vec}(\boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)})$  easily follows from Lemma A.3(ii).

Finally, we consider case (iv), under which  $\delta_n = o(1/\sqrt{n})$  and  $v_n = O(1)$ . It directly follows from (22) and (24) that  $L_{n1}$  and  $L_{n2}$  are then  $o_P(1)$ , so that  $\Lambda_n$  also is.  $\square$

## Appendix B: Proofs of Theorems 3.1–3.2

The proofs of this section require the following preliminary result.

**Lemma B.1.** *Let  $(\mathbf{V}_{0n})$  be a sequence of null shape matrices as in (12) and denote Tyler’s M-estimator of scatter as  $\hat{\mathbf{V}}_n$ . Then, we have the following under  $P_{\mathbf{V}_{0n}}^{(n)}$  as  $n \rightarrow \infty$ :*

(i) letting  $\mathbf{G}_p := \mathbf{I}_{p^2} - \frac{1}{p+2}(\mathbf{I}_{p^2} + \mathbf{K}_p - \mathbf{J}_p)$ ,

$$\mathbf{G}_p(\mathbf{V}_{0n}^{-1/2} \otimes \mathbf{V}_{0n}^{-1/2})\sqrt{n} \text{vec}(\hat{\mathbf{V}}_n - \mathbf{V}_{0n}) = p\sqrt{n} \text{vec}\left(\mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p}\mathbf{I}_p\right) + o_P(1);$$

(ii)  $\sqrt{n}(\hat{\mathbf{V}}_n - \mathbf{V}_{0n})$  is  $O_P(1)$ .

**Proof of Lemma B.1.** . Part (i) of the lemma follows from (3.7)–(3.8) in Tyler [37] and Lemma A.3(i), whereas Part (ii) follows from Part (i) and Lemma A.3(ii).  $\square$

In the proofs of this section, all stochastic convergences will be as  $n \rightarrow \infty$  under  $P_{\mathbf{V}_{0n}}^{(n)}$ .

**Proof of Theorem 3.1.** Standard properties of the vec operator provide

$$\begin{aligned}
 & \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{vec} \left( \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni}} - \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \right) \\
 &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni} - \mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}}{(\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni})(\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni})} \text{vec}(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}) \\
 &= -\left( \frac{1}{n} \sum_{i=1}^n \frac{\text{vec}(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2})(\text{vec}(\mathbf{U}_{ni} \mathbf{U}'_{ni}))'}{(\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni})(\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni})} \right) \sqrt{n} \text{vec}(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \\
 &= -(\mathbf{M}_{1n} + \mathbf{M}_{2n}) \sqrt{n} \text{vec}(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}),
 \end{aligned}$$

where

$$\begin{aligned}
 \mathbf{M}_{1n} &:= \frac{1}{n} \sum_{i=1}^n \frac{\text{vec}(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2})(\text{vec}(\mathbf{U}_{ni} \mathbf{U}'_{ni}))'}{(\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni})^2} \\
 &= ((\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2}) \otimes (\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2})) \mathbf{T}_n(\mathbf{V}_{0n})(\mathbf{V}_{0n}^{1/2} \otimes \mathbf{V}_{0n}^{1/2})
 \end{aligned}$$

and

$$\begin{aligned}
 \mathbf{M}_{2n} &:= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni}} - \frac{1}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \right\} \frac{\text{vec}(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2})(\text{vec}(\mathbf{U}_{ni} \mathbf{U}'_{ni}))'}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \\
 &= -\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{U}'_{ni} (\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \mathbf{U}_{ni}}{(\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni})(\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni})^2} \text{vec}(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2})(\text{vec}(\mathbf{U}_{ni} \mathbf{U}'_{ni}))'.
 \end{aligned}$$

The squared Frobenius norm  $\|\mathbf{M}_{2n}\|_F^2 = \text{tr}[\mathbf{M}_{2n} \mathbf{M}'_{2n}]$  is

$$\begin{aligned}
 \|\mathbf{M}_{2n}\|_F^2 &= \frac{1}{n^2} \sum_{i,j=1}^n \frac{(\mathbf{U}'_{ni} (\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \mathbf{U}_{ni})(\mathbf{U}'_{nj} (\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \mathbf{U}_{nj})}{(\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni})(\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni})^2 (\mathbf{U}'_{nj} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{nj})(\mathbf{U}'_{nj} \mathbf{V}_{0n}^{-1} \mathbf{U}_{nj})^2} \\
 &\quad \times (\mathbf{U}'_{ni} \mathbf{U}_{nj})^2 (\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{nj})^2,
 \end{aligned}$$

so that, denoting as  $\rho(\mathbf{A})$  the spectral radius of  $\mathbf{A}$ , the Cauchy–Schwarz inequality yields

$$\begin{aligned}
 \|\mathbf{M}_{2n}\|_F^2 &\leq \frac{1}{n^2} \sum_{i,j=1}^n \frac{|\mathbf{U}'_{ni} (\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \mathbf{U}_{ni}| |\mathbf{U}'_{nj} (\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \mathbf{U}_{nj}|}{(\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni})^2 (\mathbf{U}'_{nj} \mathbf{V}_{0n}^{-1} \mathbf{U}_{nj})^2} \\
 &\leq \frac{\sup_{\mathbf{u} \in \mathcal{S}^{p-1}} (\mathbf{u}' (\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \mathbf{u})^2}{\inf_{\mathbf{u} \in \mathcal{S}^{p-1}} (\mathbf{u}' \mathbf{V}_{0n}^{-1} \mathbf{u})^4} = \lambda_{n1}^4 (\rho(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}))^2,
 \end{aligned}$$

which is  $o_P(1)$  (note indeed that Lemma B.1(ii) implies that  $\sqrt{n}(\tilde{\mathbf{V}}_{0n} - \mathbf{V}_{0n})$ , hence also  $\sqrt{n}(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1})$ , is  $O_P(1)$ ). Consequently, we have proved that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{vec} \left( \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni}} - \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \right) \\ &= -((\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2}) \otimes (\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2})) \mathbf{T}_n(\mathbf{V}_{0n})(\mathbf{V}_{0n}^{1/2} \otimes \mathbf{V}_{0n}^{1/2}) \sqrt{n} \text{vec}(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) + o_P(1). \end{aligned}$$

Still using the fact that  $\sqrt{n}(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1})$  is  $O_P(1)$ , we then obtain from Lemma A.3(i) and the continuous mapping theorem that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{vec} \left( \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni}} - \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \right) \\ &= -\frac{1}{p(p+2)} ((\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2}) \otimes (\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2})) (\mathbf{I}_p + \mathbf{K}_p + \mathbf{J}_p) \\ &\quad \times (\mathbf{V}_{0n}^{1/2} \otimes \mathbf{V}_{0n}^{1/2}) \sqrt{n} \text{vec}(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) + o_P(1). \end{aligned} \quad (29)$$

Now, since  $\tilde{\mathbf{V}}_{0n}^{-1/2} - \mathbf{V}_{0n}^{-1/2}$  is  $O_P(1/\sqrt{n})$ , Lemma A.3(ii) provides

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{vec} \left( \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} - \frac{\mathbf{V}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \right) \\ &= \sqrt{n} [(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2} \otimes \tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2}) - \mathbf{I}_{p^2}] \text{vec}(\mathbf{S}(\mathbf{V}_{0n})) \\ &= \frac{\sqrt{n}}{p} [(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2} \otimes \tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2}) - \mathbf{I}_{p^2}] \text{vec}(\mathbf{I}_p) + o_P(1). \end{aligned} \quad (30)$$

Therefore, (29)–(30) provide

$$\begin{aligned} & \sqrt{n} \text{vec}(\mathbf{S}_n(\tilde{\mathbf{V}}_{0n}) - \mathbf{S}_n(\mathbf{V}_{0n})) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{vec} \left( \frac{\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \tilde{\mathbf{V}}_{0n}^{-1} \mathbf{U}_{ni}} - \frac{\mathbf{V}_{0n}^{-1/2} \mathbf{U}_{ni} \mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1/2}}{\mathbf{U}'_{ni} \mathbf{V}_{0n}^{-1} \mathbf{U}_{ni}} \right) \\ &= -\frac{1}{p(p+2)} (\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2} \otimes \tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2}) (\mathbf{I}_p + \mathbf{K}_p + \mathbf{J}_p) (\mathbf{V}_{0n}^{1/2} \otimes \mathbf{V}_{0n}^{1/2}) \sqrt{n} \text{vec}(\tilde{\mathbf{V}}_{0n}^{-1} - \mathbf{V}_{0n}^{-1}) \\ &\quad + \frac{\sqrt{n}}{p} [(\tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2} \otimes \tilde{\mathbf{V}}_{0n}^{-1/2} \mathbf{V}_{0n}^{1/2}) - \mathbf{I}_{p^2}] \text{vec}(\mathbf{I}_p) + o_P(1). \end{aligned}$$

Hence, using the fact that  $\theta_0$  is an eigenvector of any matrix power of  $\tilde{\mathbf{V}}_{0n}$  and  $\mathbf{V}_{0n}$  along with the identity  $(\mathbf{I}_p - \theta_0 \theta_0') \theta_0 = \mathbf{0}$ , we then obtain

$$(\theta_0' \otimes (\mathbf{I}_p - \theta_0 \theta_0')) \sqrt{n} \text{vec}(\mathbf{S}_n(\tilde{\mathbf{V}}_{0n}) - \mathbf{S}_n(\mathbf{V}_{0n})) = o_P(1),$$

which finally proves that

$$T_n(\tilde{\mathbf{V}}_{0n}) - T_n(\mathbf{V}_{0n}) = np(p+2) \left\| (\boldsymbol{\theta}'_0 \otimes (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0)) \text{vec}(\mathbf{S}_n(\tilde{\mathbf{V}}_{0n}) - \mathbf{S}_n(\mathbf{V}_{0n})) \right\|^2.$$

is  $o_P(1)$ . □

The proof of Theorem 3.2 still requires the following result.

**Lemma B.2.** *Let  $(\mathbf{V}_{0n})$  be a sequence of null shape matrices as in (12) and assume that there exists  $\eta > 0$  such that both corresponding leading eigenvalues satisfy  $\lambda_{n1}/\lambda_{n2} \geq 1 + \eta$  for  $n$  large enough. Then, as  $n \rightarrow \infty$  under  $P_{\mathbf{V}_{0n}}^{(n)}$ ,*

$$\frac{1}{\hat{\lambda}_{n1}} \sum_{j=2}^p \hat{\lambda}_{nj}^{-1} \hat{\boldsymbol{\theta}}_{nj} \hat{\boldsymbol{\theta}}'_{nj} = \frac{1}{\lambda_{n1}} \sum_{j=2}^p \lambda_{nj}^{-1} \boldsymbol{\theta}_{nj} \boldsymbol{\theta}'_{nj} + o_P(1),$$

where the  $\hat{\lambda}_{nj}$ 's and  $\hat{\boldsymbol{\theta}}_{nj}$ 's refer to the spectral decomposition of Tyler's M-estimator of scatter as  $\hat{\mathbf{V}}_n$  in (13).

**Proof of Lemma B.2.** It follows from Lemma B.1(ii) that  $\hat{\mathbf{V}}_n^{-1} - \mathbf{V}_{0n}^{-1}$  is  $o_P(1)$  as  $n \rightarrow \infty$  under  $P_{\mathbf{V}_{0n}}^{(n)}$ . Since  $\lambda_{n1}/\lambda_{n2}$  stays away from one, we also have that  $\hat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_{n1}$  is  $o_P(1)$  under the same sequence of hypotheses. Consequently,

$$(\mathbf{I}_p - \hat{\boldsymbol{\theta}}_{n1} \hat{\boldsymbol{\theta}}'_{n1}) \hat{\mathbf{V}}_n^{-1} - (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \mathbf{V}_{0n}^{-1} = \sum_{j=2}^p \hat{\lambda}_{nj}^{-1} \hat{\boldsymbol{\theta}}_{nj} \hat{\boldsymbol{\theta}}'_{nj} - \sum_{j=2}^p \lambda_{nj}^{-1} \boldsymbol{\theta}_{nj} \boldsymbol{\theta}'_{nj}$$

is  $o_P(1)$  as  $n \rightarrow \infty$  under  $P_{\mathbf{V}_{0n}}^{(n)}$ . The result then follows from the fact that Lemma B.1(ii) also implies that  $\hat{\lambda}_{n1} - \lambda_{n1}$  is  $o_P(1)$  as  $n \rightarrow \infty$  under  $P_{\mathbf{V}_{0n}}^{(n)}$ . □

**Proof of Theorem 3.2.** Since  $\tilde{\mathbf{V}}_{0n} \boldsymbol{\theta}_0 = \hat{\lambda}_{n1} \boldsymbol{\theta}_0$ , we have that

$$\begin{aligned} L_n &= \frac{np}{p+2} (\hat{\lambda}_{n1} \boldsymbol{\theta}'_0 \hat{\mathbf{V}}_n^{-1} \boldsymbol{\theta}_0 + \hat{\lambda}_{n1}^{-1} \boldsymbol{\theta}'_0 \hat{\mathbf{V}}_n \boldsymbol{\theta}_0 - 2) \\ &= \frac{np}{p+2} \sum_{j=1}^p (\hat{\lambda}_{n1} \hat{\lambda}_{nj}^{-1} (\boldsymbol{\theta}'_0 \hat{\boldsymbol{\theta}}_{nj})^2 + \hat{\lambda}_{n1}^{-1} \hat{\lambda}_{nj} (\boldsymbol{\theta}'_0 \hat{\boldsymbol{\theta}}_{nj})^2 - 2(\boldsymbol{\theta}'_0 \hat{\boldsymbol{\theta}}_{nj})^2) \\ &= \frac{np}{(p+2) \hat{\lambda}_{n1}} \sum_{j=2}^p \hat{\lambda}_{nj}^{-1} \{(\hat{\lambda}_{nj} - \hat{\lambda}_{n1})(\boldsymbol{\theta}'_0 \hat{\boldsymbol{\theta}}_{nj})\}^2 \\ &= \frac{np}{(p+2) \hat{\lambda}_{n1}} \sum_{j=2}^p \hat{\lambda}_{nj}^{-1} \{\hat{\boldsymbol{\theta}}'_{nj} (\hat{\mathbf{V}}_n - \tilde{\mathbf{V}}_{0n}) \boldsymbol{\theta}_0\}^2. \end{aligned}$$

Since Lemma B.1(ii) entails that  $\sqrt{n}(\hat{\mathbf{V}}_n - \tilde{\mathbf{V}}_{0n})$  is  $O_P(1)$  and since both  $\tilde{\mathbf{V}}_{0n}\boldsymbol{\theta}_0 = \hat{\lambda}_{n1}\boldsymbol{\theta}_0$  and  $\mathbf{V}_{0n}\boldsymbol{\theta}_0 = \lambda_{n1}\boldsymbol{\theta}_0$  are orthogonal to  $\boldsymbol{\theta}_{nj}$ ,  $j = 2, \dots, n$ , Lemma B.2 then provides

$$\begin{aligned} L_n &= \frac{np}{(p+2)\lambda_{n1}} \sum_{j=2}^p \lambda_{nj}^{-1} (\boldsymbol{\theta}'_{nj} (\hat{\mathbf{V}}_n - \tilde{\mathbf{V}}_{0n}) \boldsymbol{\theta}_0)^2 + o_P(1) \\ &= \frac{np}{(p+2)\lambda_{n1}} \sum_{j=2}^p \lambda_{nj}^{-1} (\boldsymbol{\theta}'_{nj} (\hat{\mathbf{V}}_n - \mathbf{V}_{0n}) \boldsymbol{\theta}_0)^2 + o_P(1), \end{aligned}$$

which, letting  $\mathbf{G}_p := \mathbf{I}_{p^2} - (1/(p+2))(\mathbf{I}_{p^2} + \mathbf{K}_p - \mathbf{J}_p)$ , rewrites

$$\begin{aligned} L_n &= \frac{p+2}{p} \sum_{j=2}^p \left\{ \frac{p}{(p+2)\sqrt{\lambda_{n1}\lambda_{nj}}} (\boldsymbol{\theta}'_0 \otimes \boldsymbol{\theta}'_{nj}) \sqrt{n} \text{vec}(\hat{\mathbf{V}}_n - \mathbf{V}_{0n}) \right\}^2 + o_P(1) \\ &= \frac{p+2}{p} \sum_{j=2}^p \left\{ (\boldsymbol{\theta}'_0 \otimes \boldsymbol{\theta}'_{nj}) \mathbf{G}_p (\mathbf{V}_{0n}^{-1/2} \otimes \mathbf{V}_{0n}^{-1/2}) \sqrt{n} \text{vec}(\hat{\mathbf{V}}_n - \mathbf{V}_{0n}) \right\}^2 + o_P(1). \end{aligned}$$

By using Lemma B.1(i), we therefore conclude that

$$\begin{aligned} L_n &= p(p+2) \sum_{j=2}^p \left\{ (\boldsymbol{\theta}'_0 \otimes \boldsymbol{\theta}'_{nj}) \sqrt{n} \text{vec} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \right\}^2 + o_P(1) \\ &= np(p+2) \left( \text{vec} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \right)' \left( (\boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \otimes (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0) \right) \text{vec} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \\ &\quad + o_P(1) \\ &= T_n(\mathbf{V}_{0n}) + o_P(1), \end{aligned}$$

which, in view of Theorem 3.1, establishes the result.  $\square$

## Appendix C: Proofs of Theorems 4.1–4.2

**Proof of Theorem 4.1.** It directly follows from Lemma A.3 that, under  $P_{\mathbf{V}_{0n}}^{(n)}$ ,

$$\begin{aligned} \mathbf{W}_n &:= (\boldsymbol{\theta}'_0 \otimes (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0)) \sqrt{n} \text{vec}(\mathbf{S}_n(\mathbf{V}_{0n})) \\ &= (\boldsymbol{\theta}'_0 \otimes (\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0)) \sqrt{n} \text{vec} \left( \mathbf{S}_n(\mathbf{V}_{0n}) - \frac{1}{p} \mathbf{I}_p \right) \end{aligned}$$

is asymptotically normal with mean zero and covariance matrix  $\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0$ . Since  $\mathbf{I}_p - \boldsymbol{\theta}_0 \boldsymbol{\theta}'_0$  is idempotent with rank  $p-1$ , this implies that, under the same sequence of hypotheses,  $\mathbf{T}_n(\mathbf{V}_{0n}) =$

$\mathbf{W}'_n(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0)\mathbf{W}_n$  is asymptotically chi-square with  $p - 1$  degrees of freedom. The result then follows from Theorem 3.1. □

**Proof of Theorem 4.2.** The results in (i)–(ii) follow from a routine application of the Le Cam third lemma. For (iii), the mutual contiguity between  $P_{\mathbf{V}_{0n}}^{(n)}$  and  $P_{\mathbf{V}_{1n}}^{(n)}$  – which follows by applying the Le Cam first lemma to Theorem 2.1(iii) – enables the use of the same Le Cam third lemma. Using the notation introduced in the proof of Theorem 4.1, the central limit theorem yields that, under  $P_{\mathbf{V}_{0n}}^{(n)}$ ,

$$\left( \begin{array}{c} \mathbf{W}_n \\ \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\theta}_0 + \frac{1}{2\xi} \boldsymbol{\tau}'_n \boldsymbol{\Upsilon}_{\boldsymbol{\theta}_0}^{(n)} \boldsymbol{\tau}_n \end{array} \right)$$

is asymptotically Gaussian with mean zero and covariance matrix

$$\left( \begin{array}{cc} \mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0 & \frac{\sqrt{p}}{\sqrt{p+2}} \left( 1 - \frac{1}{2\xi^2} \|\boldsymbol{\tau}\|^2 \right) (\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0) \boldsymbol{\tau} \\ \frac{\sqrt{p}}{\sqrt{p+2}} \left( 1 - \frac{1}{2\xi^2} \|\boldsymbol{\tau}\|^2 \right) \boldsymbol{\tau}' (\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0) & \frac{p}{p+2} \|\boldsymbol{\tau}_n\|^2 \left( 1 - \frac{1}{4\xi^2} \|\boldsymbol{\tau}_n\|^2 \right) \end{array} \right).$$

The Le Cam third lemma therefore ensures that, under the sequence of local alternatives considered in Part (iii) of the theorem,  $\mathbf{W}_n$  is asymptotically normal with mean  $\frac{\sqrt{p}}{\sqrt{p+2}}(1 - \frac{1}{2\xi^2} \|\boldsymbol{\tau}\|^2)(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0)\boldsymbol{\tau}$  and covariance matrix  $\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0$ . It follows that  $\mathbf{T}_n(\mathbf{V}_{0n}) = \mathbf{W}'_n(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}'_0)\mathbf{W}_n$  is asymptotically chi-square with  $p - 1$  degrees of freedom and non-centrality parameter

$$\frac{p}{p+2} \|\boldsymbol{\tau}\|^2 \left( 1 - \frac{1}{2\xi^2} \|\boldsymbol{\tau}\|^2 \right)^2 \left( 1 - \frac{1}{4\xi^2} \|\boldsymbol{\tau}\|^2 \right).$$

From contiguity, Theorem 3.1 implies that the same holds for  $T_n(\tilde{\mathbf{V}}_{0n})$ , which establishes the result. Finally, Part (iv) of the result directly follows from Theorem 2.1(iv). □

### Acknowledgements

Davy Paindaveine’s research is supported by a research fellowship from the Francqui Foundation and by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles. Thomas Verdebout’s research is supported by the Crédit de Recherche J.0134.18 of the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique, and by the aforementioned ARC program of the Université libre de Bruxelles.

### References

[1] Anderson, T.W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 <https://doi.org/10.1214/aoms/1177704248>

- [2] Antoine, B. and Lavergne, P. (2014). Conditional moment models under semi-strong identification. *J. Econometrics* **182** 59–69. MR3212761 <https://doi.org/10.1016/j.jeconom.2014.04.008>
- [3] Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *J. Econometrics* **133** 443–477. MR2252905 <https://doi.org/10.1016/j.jeconom.2005.06.007>
- [4] Dümbgen, L. (1998). On Tyler’s  $M$ -functional of scatter in high dimension. *Ann. Inst. Statist. Math.* **50** 471–491. MR1664575 <https://doi.org/10.1023/A:1003573311481>
- [5] Dürre, A., Fried, R. and Vogel, D. (2015). The spatial sign covariance matrix and its application for robust correlation estimation. *Aust. J. Stat.* **46** 13–22.
- [6] Dürre, A., Tyler, D.E. and Vogel, D. (2016). On the eigenvalues of the spatial sign covariance matrix in more than two dimensions. *Statist. Probab. Lett.* **111** 80–85. MR3474786 <https://doi.org/10.1016/j.spl.2016.01.009>
- [7] Dürre, A., Vogel, D. and Fried, R. (2015). Spatial sign correlation. *J. Multivariate Anal.* **135** 89–105. MR3306428 <https://doi.org/10.1016/j.jmva.2014.12.002>
- [8] Flury, B. (1988). *Common Principal Components and Related Multivariate Models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: Wiley. MR0986245 [https://doi.org/10.1007/978-3-642-01932-6\\_28](https://doi.org/10.1007/978-3-642-01932-6_28)
- [9] Forchini, G. and Hillier, G. (2003). Conditional inference for possibly unidentified structural equations. *Econometric Theory* **19** 707–743. MR2002577 <https://doi.org/10.1017/S0266466603195011>
- [10] Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170 <https://doi.org/10.1214/aos/1031689019>
- [11] Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465 <https://doi.org/10.1214/009053606000000731>
- [12] Hallin, M., Paindaveine, D. and Verdebout, T. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38** 3245–3299. MR2766852 <https://doi.org/10.1214/10-AOS810>
- [13] Hallin, M., Paindaveine, D. and Verdebout, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli* **19** 2524–2556. MR3160563 <https://doi.org/10.3150/12-BEJ461>
- [14] Hettmansperger, T.P. and Randles, R.H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860. MR1946515 <https://doi.org/10.1093/biomet/89.4.851>
- [15] Jolicoeur, P. (1984). Principal components, factor analysis, and multivariate allometry: A small-sample direction test. *Biometrics* **40** 685–690.
- [16] Ley, C. and Verdebout, T. (2017). *Modern Directional Statistics*. Chapman & Hall/CRC Interdisciplinary Statistics Series. Boca Raton, FL: CRC Press. MR3752655
- [17] Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.* **31** 807–832. MR1994731 <https://doi.org/10.1214/aos/1056562463>
- [18] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*. Wiley Series in Probability and Statistics. Chichester: Wiley. MR1828667
- [19] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213. MR1346895 <https://doi.org/10.1080/10485259508832643>
- [20] Murray, P.M., Browne, R.P. and McNicholas, P.D. (2016). *uskewFactors*: Model-based clustering via mixtures of unrestricted skew-t sactor analyzer models. R package. Available at <https://cran.r-project.org/web/packages/uskewFactors/index.html>.
- [21] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Lecture Notes in Statistics **199**. New York: Springer. MR2598854 <https://doi.org/10.1007/978-1-4419-0468-3>

- [22] Paindaveine, D. (2006). A Chernoff–Savage result for shape: On the non-admissibility of pseudo-Gaussian methods. *J. Multivariate Anal.* **97** 2206–2220. MR2301635 <https://doi.org/10.1016/j.jmva.2005.08.005>
- [23] Paindaveine, D. (2009). On multivariate runs tests for randomness. *J. Amer. Statist. Assoc.* **104** 1525–1538. MR2597001 <https://doi.org/10.1198/jasa.2009.tm09047>
- [24] Paindaveine, D., Remy, J. and Verdebout, T. (2020). Testing for principal component directions under weak identifiability. *Ann. Statist.* **48** 324–345. MR4065164 <https://doi.org/10.1214/18-AOS1805>
- [25] Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli* **22** 1745–1769. MR3474832 <https://doi.org/10.3150/15-BEJ710>
- [26] Pötscher, B.M. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* **70** 1035–1065. MR1910411 <https://doi.org/10.1111/1468-0262.00318>
- [27] Randles, R.H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045–1050. MR1134492
- [28] Randles, R.H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.* **95** 1263–1268. MR1792189 <https://doi.org/10.2307/2669766>
- [29] Salibián-Barrera, M., Van Aelst, S. and Willems, G. (2006). Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. *J. Amer. Statist. Assoc.* **101** 1198–1211. MR2328307 <https://doi.org/10.1198/016214506000000096>
- [30] Schwartzman, A., Mascarenhas, W.F. and Taylor, J.E. (2008). Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann. Statist.* **36** 2886–2919. MR2485016 <https://doi.org/10.1214/08-AOS628>
- [31] Taskinen, S., Kankainen, A. and Oja, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21. MR1965367 [https://doi.org/10.1016/S0167-7152\(02\)00399-1](https://doi.org/10.1016/S0167-7152(02)00399-1)
- [32] Taskinen, S., Koch, I. and Oja, H. (2012). Robustifying principal component analysis with spatial sign vectors. *Statist. Probab. Lett.* **82** 765–774. MR2899518 <https://doi.org/10.1016/j.spl.2012.01.001>
- [33] Taskinen, S., Oja, H. and Randles, R.H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.* **100** 916–925. MR2201019 <https://doi.org/10.1198/016214505000000097>
- [34] Tyler, D.E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736. MR0619278
- [35] Tyler, D.E. (1983). A class of asymptotic tests for principal component vectors. *Ann. Statist.* **11** 1243–1250. MR0720269 <https://doi.org/10.1214/aos/1176346337>
- [36] Tyler, D.E. (1983). The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Statist.* **11** 1232–1242. MR0720268 <https://doi.org/10.1214/aos/1176346336>
- [37] Tyler, D.E. (1987). A distribution-free  $M$ -estimator of multivariate scatter. *Ann. Statist.* **15** 234–251. MR0885734 <https://doi.org/10.1214/aos/1176350263>
- [38] Tyler, D.E. (1987). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika* **74** 579–589. MR0909362 <https://doi.org/10.1093/biomet/74.3.579>
- [39] Zhu, H. and Zhang, H. (2006). Asymptotics for estimation and testing procedures under loss of identifiability. *J. Multivariate Anal.* **97** 19–45. MR2208842 <https://doi.org/10.1016/j.jmva.2004.11.005>

Received December 2018 and revised August 2019