Testing for sphericity using spatial signs under elliptical directions

Gaspard Bernard[1,2],

[1] *Department of Mathematics, University of Luxembourg, e-mail:* gaspard.bernard@uni.lu

[2] *Institute of Statistical Science, Academia Sinica*

Abstract: In this contribution, we consider the problem of testing for the sphericity of a collection of random vectors. It is well known that in a classical elliptical model, testing for rotational symmetry of the underlying distribution is equivalent to testing that a scatter parameter is a multiple of the identity matrix. We consider the more general model of random vectors with elliptical directions and introduce a few scenarios where testing for sphericity is still equivalent to testing that the scatter parameter is a multiple of the identity. These new scenarios include, for instance, non-classical settings where some dependence of a very general form studied here for the first time may be present between observations. We study, under these new assumptions, the behavior of the classical spatial sign test and show that under certain mild assumptions, the test is asymptotically valid and has the same local asymptotic power as in the classical elliptical scenario. We then show that, contrary to some commonly held belief, the spatial sign test enjoys some local asymptotic optimality properties when it comes to testing for sphericity when the underlying distribution is strongly heavy-tailed.

MSC2020 subject classifications: Primary 62H15, 62F35; secondary 62F05.
Keywords and phrases: Robust hypothesis tests, Local asymptotic optimality, Spatial signs, Shape matrices.

1. Introduction

A p-dimensional random vector $\mathbf{X}$ is considered to be *spherical* if there exists a vector $\boldsymbol{\theta} \in \mathbb{R}^p$ such that $\mathbf{X} - \boldsymbol{\theta} \overset{\mathcal{L}}{=} \mathbf{O}(\mathbf{X} - \boldsymbol{\theta})$, for all $\mathbf{O} \in \mathcal{SO}_p$, with $\mathcal{SO}_p$ the special orthogonal group of $p \times p$ matrices. As spherical random vectors do not admit a direction in which the variability is larger, they can be considered as pure noise in a large amount of applications. Sphericity tests are therefore particularly useful for trying to detect the presence of a signal in a given data set, this detection being a preliminary step to performing any multivariate analysis technique. In particular, testing for sphericity is a crucial step that must be carried out before performing Principal Component Analysis (PCA) or, more generally, before applying a dimension reduction technique to the data. For this reason, the problem of testing for sphericity has been the subject of extensive study for a very long time. [14] was the first attempt to tackle the problem by proposing a Gaussian Likelihood Ratio Test (LRT), while a locally most powerful invariant Gaussian test was obtained by [9] (see also [10]). Since the above tests are asymptotically valid only under Gaussian assumption, pseudo-Gaussian procedures requiring only finite fourth-order moments were introduced by [16] and [26]. To obtain robustness to the absence of finite fourth-order moments, a sign-based test statistic was proposed by [27], while [6] proposed tests based on the multivariate signed-ranks of the observations. Although the problem of testing for sphericity has been considered for a very long time, it is still actively studied because of its crucial importance in dimension reduction. To cite only a few recent contributions, [8] proposed tests based on characteristic functions, [2] tests based on random projections, [4] studied sphericity tests in a time series framework and [23] proposed an approach based on optimal transport. High-dimensional sphericity tests have also been extensively studied these past years; see for instance [18, 19, 25, 21] and [13].

When testing for sphericity, certain assumptions over the underlying distribution allow us to reformulate the problem in such a way as to obtain asymptotic optimality properties. If we assume that $\mathbf{X}$ is an *elliptical* random vector, the geometry of the probability contours is entirely determined by a matrix acting as a dispersion parameter. This (normalized) $p \times p$ matrix is called the *shape parameter* and we will denote it $\mathbf{V}$. The matrix $\mathbf{V}$ is normalized so that it has determinant 1, as this choice is canonical in a sense established in [20]. In the elliptical framework, the sphericity assumption is equivalent to assuming that $\mathbf{V} = \mathbf{I}_p$. If we do not assume that the random vectors are elliptical, the latter statement is not true in general and $\mathbf{V} = \mathbf{I}_p$ does not necessarily imply sphericity. For these reasons, the locally and asymptotically optimal tests proposed in [6] assume ellipticity; testing for sphericity then boiling down to testing that $\mathbf{V} = \mathbf{I}_p$. However, there may be less stringent assumptions that could be made on the data generating process under which testing for $\mathbf{V} = \mathbf{I}_p$ is still equivalent to testing for sphericity. The first objective of this contribution is to propose a model, less stringent than the classical elliptical one, in which the hypothesis $\mathbf{V} = \mathbf{I}_p$ is still equivalent to the sphericity hypothesis, making of the testing problem a (semi-)parametric one. The second objective is to propose tests that are asymptotically valid under these new assumptions and study their local asymptotic power. The third objective is to show under which conditions these tests enjoy some asymptotic optimality properties. As we shall see, our approach will lead to the use of procedures based on the *spatial signs* of the observations, in the spirit of [27]. These procedures are based solely on the directions of the (centred) observations, directions belonging to the unit sphere of dimension $p - 1$. This contribution can also be seen as a strong plea for the use of multivariate sign tests since, as we shall see later, they enjoy some highly desirable properties. In this sense, this contribution investigates an interesting phenomenon in statistical inference by pointing out that there are cases where using *less* seemingly relevant information from the data not only increases robustness but is also *optimal*.

This contribution is structured as follows. The section 2 presents the model considered, the section 3 presents the test considered and outlines our approach, the sections 4 and 5 examine the asymptotic validity, power and optimality of the test considered, while the section 6 presents our conclusions. An Appendix gathers the technical proofs of the different results. The main tool used in this contribution is the *Le Cam asymptotic theory of experiments* developed in [11] and [12]. Our results are illustrated by Monte-Carlo simulation studies.

2. The model: skewness and dependent observations

The class of *distributions with elliptical directions* was introduced in [22], generalizing the classical class of *elliptical distributions*. In the p-variate framework, distributions with elliptical directions are characterized by a *location* parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, and a *shape* parameter $\mathbf{V}$ belonging to the set of $p \times p$ positive definite matrices with determinant equal to one. Some nonparametric nuisance of a very general form is also present in the model. Explicitly, a collection of p-variate random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ follows a distribution with elliptical directions with parameters $\boldsymbol{\theta}$ and $\mathbf{V}$ if it satisfies

$$\mathbf{X}_i = R_i \mathbf{V}^{1/2} \mathbf{U}_i + \boldsymbol{\theta},\ i \in \{1, \ldots, n\}. \tag{1}$$

Here, $\mathbf{U}_1, \ldots, \mathbf{U}_n$ are i.i.d. uniform random vectors on the unit sphere

$$\mathcal{S}^{p-1} := \{\mathbf{u} \in \mathbb{R}^p \mid \mathbf{u}'\mathbf{u} = 1\}$$

and $R_1, \ldots, R_n$ is a collection of random variables with values in $\mathbb{R}^+$. We denote this model by $\mathrm{P}^{(n)}_{\boldsymbol{\theta}, \mathbf{V}, R_1, \ldots, R_n}$. The main difference with classical elliptical random vectors lies in the fact that $R_1, \ldots, R_n$ can be dependent on the $\mathbf{U}_i$, can be dependent on each others or can be such that they are not identically distributed.

In this context, the shape parameter $\mathbf{V}$ summarizes at least some part of the dispersion of the data. We distinguish three settings which are of particular relevance.

- (I) The R_i are independent of the $\mathbf{U}_i$ and i.i.d. This is the classical *elliptical* framework where the geometry of the dispersion is entirely characterized by $\mathbf{V}$.
- (II) There is some dependence between the R_i and the $\mathbf{U}_i$ but the R_i are i.i.d.. This corresponds to i.i.d. cases where some *skewness* is present in the underlying distribution. In this context, the parameter $\mathbf{V}$ does not describe the part of the geometry of the dispersion explained by the skewness.
- (III) The R_i are not i.i.d. but are independent of the $\mathbf{U}_i$. $\mathbf{V}$ still fully characterizes the geometry of the dispersion but the distribution of the radius can vary with $i = 1, \ldots, n$. In particular, *radial dependence* between the observations can be present in the R_i.

Inference over $\mathbf{V}$ in setting (I) is a very classical and well-studied problem, see for instance [24] and [6] for an approach based on signed-ranks. In this contribution, we will therefore focus mainly on settings (II) and (III), which have not been the subject of as much work. Considering inference problems on $\mathbf{V}$ in setting (II) makes sense if we assume that some contamination of a classical elliptical distribution is causing the skewness as in [5] and if we would like to conduct inference over $\mathbf{V}$ in such a way that the procedure is robust to this contamination. In general, if measurement errors are made on $R_1, \ldots, R_n$ and if these errors depend on the direction of the observations $\mathbf{U}_1, \ldots, \mathbf{U}_n$, they will most likely affect classical inference procedures performed over the dispersion parameter $\mathbf{V}$. If we suspect that such errors have been made, it is perfectly reasonable to want to perform inference on $\mathbf{V}$ in such a way that the outcome is independent of these errors. Setting (III), on the other hand, makes a lot of sense when we assume that the directions are independent, but that the overall variability evolves over time and can be correlated with previous realizations. For instance, the random vectors $\mathbf{U}_i$ could represent a direction indicating whether certain prices are rising or falling and the R_i could represent the magnitude of the price change. In this framework, the R_i can be linked to some global volatility of the market, potentially correlated to past realizations, while the directions $\mathbf{U}_i$ are still i.i.d. In setting (II) and (III), the R_i play the role of a nuisance, while the shape parameter $\mathbf{V}$ encompasses all the relevant information about the interaction between the various components of the $\mathbf{X}_i$. It is then very natural to consider classical testing problems over $\mathbf{V}$ - such as the sphericity problem - in setting (II) and (III). Furthermore, since the inference procedures based on the spatial signs of the observations are known to exhibit excellent robustness properties in general - see, for instance, [1] - it is also very natural to study the robustness of multivariate sign-based testing procedures in settings (II) and (III). We develop the rationale of our approach in the next section.

3. Testing for sphericity using multivariate signs

First, we introduce a few notations. Let $\mathbf{A}$ be a $p \times p$ matrix, we denote by $\mathrm{vec}(\mathbf{A})$ the p^2-vector obtained by stacking the columns of $\mathbf{A}$ one on top of the other and by $\overset{\circ}{\mathrm{vech}}\,(\mathbf{A})$ the vector obtained by stacking the upper-diagonal entries of the matrix $\mathbf{A}$ deprived of its first component. We also denote by $\mathbf{A}^-$ the Moore-Penrose generalized inverse of $\mathbf{A}$ and by $\mathbf{K}_p$ the $p^2 \times p^2$ *commutation matrix* defined such that $\mathbf{K}_p \mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\mathbf{A}')$. Finally, let $\mathbf{M}_p(\mathbf{V})$ the $(p(p+1)/2-1) \times p^2$ matrix such that (i) $\mathbf{M}_p(\mathbf{V})\mathrm{vec}\,(\mathbf{V}^{-1}) = \mathbf{0}$ and (ii) $\mathbf{M}_p(\mathbf{V})'\overset{\circ}{\mathrm{vech}}\,(\mathbf{A}) = \mathrm{vec}\,(\mathbf{A})$ for any symmetric matrix $\mathbf{A}$ such that $\mathrm{tr}(\mathbf{V}^{-1}\mathbf{A}) = 0$. For the sake of brevity, we put $\mathbf{M}_p := \mathbf{M}_p(\mathbf{I}_p)$. We consider the parametrization $\boldsymbol{\delta} = (\boldsymbol{\theta}', \overset{\circ}{\mathrm{vech}}{}'(\mathbf{V}))'$. In this parametrization $\overset{\circ}{\mathrm{vech}}\,(\mathbf{V})$ is the parameter of interest and we test for $\overset{\circ}{\mathrm{vech}}\,(\mathbf{V}) \in \mathcal{H}_0$ with

$$\mathcal{H}_0 : \overset{\circ}{\mathrm{vech}}\,(\mathbf{V}) \text{ is such that } \mathbf{V} = \mathbf{I}_p. \tag{2}$$

In settings (I), (II) and (III), the interpretation of $\mathcal{H}_0$ is the same: the data consists only of spherical noise. The difference between the three settings lies only in the fact that in setting (II), the data can be considered as spherical noise with some added contamination or errors that should be ignored while in setting (III), it can be considered as correlated spherical random noise. The main idea of this contribution is to get rid of the R_i in the test statistic, since it is the only part of the model that is defined differently in the three considered settings. For that purpose, we define the *elliptical signs* of the observations

$$\mathbf{U}_i(\boldsymbol{\theta}, \mathbf{V}) = \frac{\mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})}{\|\mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})\|},$$

obviously a quantity that is measurable with respect to the $\mathbf{U}_i$. When $\mathbf{V} = \mathbf{I}_p$, we find back the classical *spatial signs*. We also define the associated *elliptical distances*

$$d_i(\boldsymbol{\theta}, \mathbf{V}) = \|\mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})\|.$$

We denote by

$$\mathbf{S}^{(n)}_{\mathrm{sign}} := \mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta}, \mathbf{I}_p) = p\, n^{-1} \sum_{i=1}^{n} \mathbf{U}_i(\boldsymbol{\theta}, \mathbf{I}_p)\mathbf{U}_i(\boldsymbol{\theta}, \mathbf{I}_p)'. \tag{3}$$

It has been well known since [6] that in setting (I), asymptotically valid multivariate sign tests for $\mathcal{H}_0$ can be based on the quantity

$$\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta}, \mathbf{I}_p) = n^{-1/2}\, \frac{1}{2}\mathbf{M}_p \sum_{i=1}^{n} \overset{\circ}{\mathrm{vech}}\left(\mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta}, \mathbf{I}_p)\right). \tag{4}$$

More precisely, for the testing problem (2), the multivariate sign test studied in [6] rejects $\mathcal{H}_0$ at the asymptotic level α when

$$\begin{aligned} S^{(n)}(\boldsymbol{\theta}) &:= (\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta}, \mathbf{I}_p))'\boldsymbol{\Gamma}^{-}_{\mathrm{sign}}(\mathbf{I}_p)\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta}, \mathbf{I}_p) = \frac{n(p+2)}{2p}\|\mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta}, \mathbf{I}_p) - \mathbf{I}_p\|^2_{\mathrm{F}} \\ &> \chi^2_{p(p+1)/2-1;1-\alpha}, \end{aligned} \tag{5}$$

where $\chi^2_{\nu;\delta}$ represents the quantile of order δ of the chi-squared distribution with ν degrees of freedom and $\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{I}_p)$ is defined as in Theorem 5.3. Note here that although the explicit forms of the matrices $\mathbf{M}_p$ and $\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{I}_p)$ are given for the sake of precision, they play no role in the problem we tackle. Furthermore, the final form of the test statistic considered in (5) is rather simple and does not involve $\mathbf{M}_p$ or $\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{I}_p)$. The two main points here are as follows:

(a) the multivariate sign test (5) based on (4) is asymptotically valid in the elliptical directions setting since (4) does not depend on the R_i;
(b) the quantity $\boldsymbol{\theta}$ still needs to be estimated in (4) to produce a useable test statistic.

A natural approach is to replace $\boldsymbol{\theta}$ in (4) by a root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$. We denote by $\phi^{(n)}_{\mathrm{sign}}$ the equivalent of the test (5) where $\boldsymbol{\theta}$ has been replace by $\hat{\boldsymbol{\theta}}^{(n)}$. In setting (I), it is shown in [6] that the estimation of $\boldsymbol{\theta}$ has no asymptotic cost. In other words, we can consider without loss of generality that we are only dealing with $\mathbf{\Delta}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{I}_p)$. This completely eliminates the dependence to the R_i in (4), asymptotically.

In the settings (II) and (III), the problem is much more difficult because the arguments used in [6] that allowed to assume that $\boldsymbol{\theta}$ was specified are no longer valid. This is why we need to study the potential cost of this estimation of $\boldsymbol{\theta}$ in (4).

Considering any root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$ of $\boldsymbol{\theta}$, the problem under study amounts to answering the following essential question: "Assuming that we are in settings (II) or (III), under what conditions on $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is it possible to assert that

$$\lim_{n\to\infty} \mathbf{\Delta}^{(n)}_{\text{sign}}(\hat{\boldsymbol{\theta}}^{(n)}, \mathbf{I}_p) \overset{\mathcal{L}}{=} \lim_{n\to\infty} \mathbf{\Delta}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{I}_p), \tag{6}$$

or, equivalently, that

$$\lim_{n\to\infty} n^{1/2}\big(\mathbf{S}^{(n)}_{\text{sign}}(\hat{\boldsymbol{\theta}}^{(n)}, \mathbf{I}_p) - \mathbf{I}_p\big) \overset{\mathcal{L}}{=} \lim_{n\to\infty} n^{1/2}\big(\mathbf{S}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{I}_p) - \mathbf{I}_p\big)\text{?''}. \tag{7}$$

If the conditions such that (7) holds are met, we can consider that

$$\mathbf{\Delta}^{(n)}_{\text{sign}}(\hat{\boldsymbol{\theta}}^{(n)}, \mathbf{I}_p)$$

asymptotically no longer depends on the R_i's anymore and is then such that its asymptotic distribution does not vary under assumptions (I), (II) and (III). In the next section, we show under which conditions (7) holds and then under which conditions the multivariate sign test for sphericity $\phi^{(n)}_{\text{sign}}$ is asymptotically valid under the elliptical directions assumption (1).

4. Asymptotic validity of the multivariate sign test under elliptical directions assumption

4.1. Asymptotic validity in presence of skewness

From now on, we assume that there exists $\hat{\boldsymbol{\theta}}^{(n)}$, a consistent estimator of $\boldsymbol{\theta}$ satisfying $n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}) = O_{\mathrm{P}}(1)$ as $n \to \infty$. We will return to the question of the existence of such an estimator at the end of this section. For any $l \in \{1, \ldots, p\}$, we denote by $[\mathbf{v}]^{(l)}$ the l-th component of $\mathbf{v}$, $\mathbf{v} \in \mathbb{R}^p$. We first tackle the case where the data generating process satisfies the assumptions of setting (II). In other words, we will study the asymptotic validity of $\phi^{(n)}_{\text{sign}}$ assuming that the R_i are i.i.d. but are not independent of the $\mathbf{U}_i$. The main tool is the following theorem.

Theorem 4.1. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ p-variate random vectors with elliptical directions satisfying assumption (II). Assume that $\mathrm{E}(R_i^{-3/2}) < \infty$. Moreover, assume that*

$$\mathrm{E}(R_i^{-1} \frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|}) = 0$$

and that for every $k, l, m \in \{1, \dots, p\}$,

$$\mathrm{E}\Big([\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|}]^{(k)}[\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|}]^{(l)}[\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|}]^{(m)}R_i^{-1}\Big) = 0.$$

Then, the following holds as $n \to \infty$,

$$\lim_{n\to\infty} n^{1/2}\big(\mathbf{S}_{\mathrm{sign}}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \mathbf{I}_p) - \mathbf{I}_p\big) \stackrel{\mathcal{L}}{=} \lim_{n\to\infty} n^{1/2}\big(\mathbf{S}_{\mathrm{sign}}^{(n)}(\boldsymbol{\theta}, \mathbf{I}_p) - \mathbf{I}_p\big).$$

A direct consequence of Theorem 4.1 is of course the following proposition, which establishes the asymptotic validity of the multivariate sign test $\phi_{\mathrm{sign}}^{(n)}$ in scenario (II), under certain conditions over the R_i that we will discuss after stating the result.

Theorem 4.2. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ p-variate random vectors with elliptical directions satisfying the assumptions of Theorem 4.1. Then, $\phi_{\mathrm{sign}}^{(n)}$ is asymptotically valid for $\mathcal{H}_0$ as $n \to \infty$ under $\mathrm{P}_{\boldsymbol{\theta},\mathbf{V},R_1,\dots,R_n}^{(n)}$.*

We need to distinguish two conditions on the R_i. The condition $\mathrm{E}(R_i^{-3/2}) < \infty$ indicates only that the distribution of the R_i must not charge 0 too much. Roughly speaking, this condition makes sense since the construction of the (empirical) spatial signs implies dividing by $\|\mathbf{X}_i - \hat{\boldsymbol{\theta}}^{(n)}\|$, a quantity which can not be too close to 0. This assumption is not particularly stringent and makes a lot of sense. The second (two-part) technical condition concerns the vector $\mathrm{E}(R_i^{-1}\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|})$ and its components and is more stringent. This condition can be interpreted as follows: skewness can be present in the distribution of the $\mathbf{X}_i$ but the R_i^{-1} and the $\mathbf{U}_i$ still need to satisfy *some* symmetry assumption; in particular, they must not be correlated. Then, the test $\phi_{\mathrm{sign}}^{(n)}$ is indeed asymptotically robust to some form of contamination or measurement errors causing skewness in the data set of the type considered in the section 2. However, very roughly speaking, the assumption that said contamination/measurement errors still preserve some form of symmetry should be met for the test $\phi_{\mathrm{sign}}^{(n)}$ to remain asymptotically valid.

4.2. Asymptotic validity in presence of radial dependence

The radial dependance scenario, where the data generating process satisfies the assumptions (III), allows to get a much more interesting result. We will now study the asymptotic consistency of $\phi_{\mathrm{sign}}^{(n)}$ assuming that the R_i are not i.i.d. but are independent of the $\mathbf{U}_i$. We have the following proposition.

Theorem 4.3. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ p-variate random vectors with elliptical directions satisfying assumption (III). Assume moreover that $R_1, \dots, R_n$ are such that $n^{-3/2}\sum_{i=1}^n R_i^{-3} = o_{\mathrm{P}}(1)$. Then, the following holds as $n \to \infty$,*

$$n^{1/2}\big(\mathbf{S}_{\mathrm{sign}}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}, \mathbf{I}_p) - \mathbf{S}_{\mathrm{sign}}^{(n)}(\boldsymbol{\theta}, \mathbf{I}_p)\big) = o_{\mathrm{P}}(1).$$

A proof can be found in the Appendix. A direct consequence of Theorem 4.3 is of course the next proposition, which establishes the asymptotic validity of $\phi^{(n)}_{\mathrm{sign}}$ in a large proportion of scenarios of type (III).

Theorem 4.4. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ p-variate random vectors with elliptical directions satisfying the assumptions of Theorem 4.3 Then, $\phi^{(n)}_{\mathrm{sign}}$ is asymptotically valid for $\mathcal{H}_0$ as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V},R_1,\dots,R_n}$.*

Theorem 4.4 is a stronger result than Theorem 4.2 in the sense that the assumption made about the R_i is not that stringent, when independence from the $\mathbf{U}_i$ is assumed. Indeed, $n^{-3/2}\sum_{i=1}^n R_i^{-3}$ acts only as a way to prevent the distribution of the R_i from charging 0 at a very fast rate and is a totally reasonable assumption that will be in practice satisfied by most of the classical examples we can think of - modulo some approximation we will detail later. As we will show, testing for sphericity using $\phi^{(n)}_{\mathrm{sign}}$ allows virtually to tackle every possible radial dependance scenario we can think of, making the multivariate sign test a very powerful tool in setting (III). We illustrate this last statement with some examples of radii $R_1, \dots, R_n$ satisfying the assumptions of Theorem 4.4.

First, we will state a very general result, allowing to model some extremely general type of radial dependence.

Theorem 4.5. *Let $\mathbf{X}_{1n}, \dots, \mathbf{X}_{nn}$ a triangular array of p-variate random vectors satisfying assumption (III) with $R_{1n} = |A_1| + c_n, \dots, R_{nn} = |A_n| + c_n$, where $(A_i)_{i\in\mathbb{Z}}$ is any arbitrary discrete stochastic process and $c_n^{-1} = o(n^{1/6})$. Then, $n^{-3/2}\sum_{i=1}^n R_i^{-3} = o_{\mathrm{P}}(1)$ as $n \to \infty$.*

A direct consequence of Theorem 4.5 is that the test (5) is asymptotically valid for any root-n consistent estimator of $\boldsymbol{\theta}$ if the radii $R_1, \dots, R_n$ are assumed to be drown in a manner that is asymptotically equivalent to any given stochastic process. Indeed, since we can have $c_n = o(1)$ (assuming that the speed of convergence of c_n^{-1} is not too fast), we have that for this choice of c_n as $n \to \infty$, $R_i = |A_i| + o_{\mathrm{P}}(1)$. Theorem 4.5 highlights the fact that the assumption of Theorem 4.4 seems perfectly reasonable and that the test $\phi^{(n)}_{\mathrm{sign}}$ can be used in most classical radial dependance scenarios we can think of. In practice, any model presenting some radial dependence can be approximated by $R_1 = |A_1| + c_n, \dots, R_n = |A_n| + c_n$ with $c_n = o(1)$ and for this approximation, the spatial sign test $\phi^{(n)}_{\mathrm{sign}}$ will always be asymptotically valid, provided that c_n does not converge too fast to 0. However, one could want to get rid of the approximation implied by considering $|A_i| + c_n$ instead of $|A_i|$, which is perfectly feasible if one requires more conditions on A_i to be satisfied.

Another very natural approach is to assume that the R_i are identically distributed but not independent. This covers, for instance, the case where for all $i = 1, \dots, n$, $\mathbf{X}_i = R_i\mathbf{V}^{1/2}\mathbf{U}_i + \boldsymbol{\theta}$ are identically distributed Gaussian p-vectors. We have the following result, which guarantees that if the dimension p is sufficiently large, the multivariate sign test is asymptotically valid in all classical elliptical identically distributed cases without having to rely on the type of approximation used in Theorem 4.5. As usual in the radial dependence scenario, we still assume that $\mathbf{U}_i$ and R_i are independent and that the $\mathbf{U}_i$ are i.i.d..

Theorem 4.6. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ identically distributed and satisfying assumption (III). Assume moreover that the $\mathbf{X}_i$ have common density $f_{\mathbf{X}_1}$. Assume moreover that $p > 6$ and that $f_{\mathbf{X}_1}$ is bounded. Under those assumptions,*

$$n^{-3/2} \sum_{i=1}^{n} R_i^{-3} = o_{\mathrm{P}}(1)$$

as $n \to \infty$.

The assumption $p > 6$ may seem rather surprising at first sight. However, as the radial part of $\mathbf{X}_i$ satisfies $R_i = \|\mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})\|$, its density depends on p via the Jacobian of the change of variable. When $p > 6$, the negative moments of order 6 are finite for most of the classical p-variate models - which means that $\mathbf{X}_i$ does not charge $\mathbf{0}$ too fast. We provide one last example of models for which testing for sphericity using spatial signs leads to asymptotically valid procedures.

Since the classical way of dealing with dependent observations is the time series framework, it makes perfect sense to consider a model where (some transformation of) $R_1, \ldots, R_n$ follow a classical time series model and assess whether the assumptions of Theorem 4.3 are satisfied. In the following proposition, we consider the log-transformation of the $R_1, \ldots, R_n$ due to both its popularity in time series analysis and technical reasons.

Theorem 4.7. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ p-variate random vectors satisfying assumption (III) with $\log(R_1) = A_1, \ldots, \log(R_n) = A_n$, where $(A_i)_{i\in\mathbb{Z}}$ is an $\mathrm{ARMA}(l, q)$ process with centred i.i.d. errors ϵ_i with finite variance.*

Moreover, assume that ϵ_1 satisfies $\prod_{i=1}^{\infty} \mathrm{E}(e^{-t_i \epsilon_1}) < \infty$ for every sequence $(t_i)_{i\in\mathbb{N}}$ satisfying $\sum_{i=1}^{\infty} t_i^2 < \infty$.

In this model, for all $i \in \mathbb{Z}$, there exist $\varphi_1, \ldots, \varphi_p < 1$ and $\theta_1, \ldots, \theta_q < 1$, such that we have

$$A_i = \sum_{k=1}^{l} \varphi_k A_{i-k} + \sum_{j=1}^{q} \theta_j \epsilon_{i-j} + \epsilon_i.$$

Under those assumptions, $n^{-3/2} \sum_{i=1}^{n} R_i^{-3} = o_{\mathrm{P}}(1)$ as $n \to \infty$.

According to Theorem 4.7, we can conclude that the test $\phi_{\mathrm{sign}}^{(n)}$ is asymptotically valid if the log-transformed radii $R_1, \ldots, R_n$ are assumed to be drown according to the very classical $\mathrm{ARMA}(l, q)$ time series model, granted that the errors ϵ_i satisfy a certain technical assumption. Although this condition may seem a little too abstract, it is not hard to check that it is satisfied in the very classical case where $\epsilon_1 \sim \mathcal{N}(0, \sigma)$, which already allows to tackle the most classical models.

The results of this section clearly show that the asymptotic robustness to radial dependence of the multivariate sign test is extremely impressive. Indeed,

$\phi^{(n)}_{\rm sign}$ is asymptotically valid under identically distributed p-variate elliptical random vectors with p sufficiently large, under classical ARMA time series models with Gaussian errors and under models approximating asymptotically *any* arbitrary stochastic process.

4.3. Estimation of the location parameter

We still need to address the question of the choice of the estimator $\hat{\boldsymbol{\theta}}^{(n)}$ of $\boldsymbol{\theta}$. We suggest using the *spatial median* $\hat{\boldsymbol{\theta}}^{(n)}_{\rm sign}$, introduced in [15]. Intuitively, this choice is logical since the spatial median is itself based on the spatial signs of the observations. The spatial median $\hat{\boldsymbol{\theta}}^{(n)}_{\rm sign}$ is defined as the p-variate random vector satisfying

$$\sum_{i=1}^{n} \mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)}_{\rm sign}, \mathbf{I}_p) = \mathbf{0}.$$

We have the following result, guaranteeing that $\hat{\boldsymbol{\theta}}^{(n)}_{\rm sign}$ is root-n consistent under some rather lenient assumptions.

Theorem 4.8. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ p-variate random vectors with elliptical directions satisfying assumptions (I), (II) or (III). Let $R_1, \ldots, R_n$ such that as $n \to \infty$,*

(i) $n^{-1} \sum_{i=1}^n {R_i}^{-1} = O_{\rm P}(1)$,
(ii) there is no subsequence $s(n)^{-1} \sum_{i=1}^{s(n)} R_i^{-1}$ *such that* $s(n)^{-1} \sum_{i=1}^{s(n)} R_i^{-1} = o_{\rm P}(1)$,
(iii) there exist $\delta \in (0,1)$ *such that* $n^{-1-\delta/2} \sum_{i=1}^n {R_i}^{-1-\delta} = o_{\rm P}(1)$.

Then, under those assumptions, $n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)}_{\rm sign} - \boldsymbol{\theta}) = O_{\rm P}(1)$ *as* $n \to \infty$.

We briefly summarize the results of this section. Thanks to Theorem 4.8, plugging the spatial median $\hat{\boldsymbol{\theta}}^{(n)}_{\rm sign}$ in the test statistic defined in (5) allows us to obtain a test for sphericity $\phi^{(n)}_{\rm sign}$ which is asymptotically robust to some skewness or to radial dependence between the observations in most cases (provided that the assumptions of Theorem 4.1 or 4.3, respectively, are satisfied). We have yet to answer the question of the asymptotic power of $\phi^{(n)}_{\rm sign}$, which is the subject of the next section but we first present the results of a Monte Carlo simulation study illustrating the results of this section.

4.4. Simulation study

We generated $M = 20,000$ independent samples $\mathbf{X}^{(j)}_1, \ldots, \mathbf{X}^{(j)}_n$ with $n = 2,000$ and $j \in \{1,2,3,4\}$. For all j, the random $(p = 3)$-variate vectors satisfy $\mathbf{X}^{(j)}_i = R^{(j)}_i \mathbf{U}_i$, $i \in \{1, \ldots, n\}$, with $\mathbf{U}_i$ uniformly distributed on $\mathcal{S}^2$.

When $j \in \{1, 2\}$, some asymmetry is present in the underlying data-generating process (scenario (II)). If $j = 1$, the $R_i^{(j)}$ satisfy the assumptions of Theorem 4.2. More precisely, $R_i^{(j)} = |Z_i| + |\langle(0, 0, 20), \mathbf{U}_i'\rangle| + 1/10$ where Z_i are i.i.d. standard Gaussian random variables with expectation 1. When $j = 2$, the $R_i^{(j)}$ do not satisfy the assumptions of Theorem 4.2 anymore since we define $R_i^{(j)} = |Z_i| + \max\{\langle(0, 0, 20), \mathbf{U}_i'\rangle, 0\} + 1/10$. We mentioned in the section 2 that inference in scenario (II) makes sense if we assume that some measurement errors were made in such a way that they depend on the directions $\mathbf{U}_1, \ldots, \mathbf{U}_n$. The simulation setups for $j \in \{1, 2\}$ can be viewed as some toy exemples where the measurement error is a simple translation of the radial component by $|\langle(0, 0, 20), \mathbf{U}_i'\rangle|$ or $\max\{\langle(0, 0, 20), \mathbf{U}_i'\rangle, 0\}$ respectively.

When $j \in \{3, 4\}$, some radial dependence is present in the data-generating process (scenario (III)). If $j = 3$, the $R_i^{(j)}$ satisfy the assumptions of Theorem 4.4: in this setup $R_i^{(j)} = |Z_i| + 1/10$ with $(Z_i \mid Z_{i-1} = z)$ following a student distribution with non-centrality parameter z, and 2.9 degrees of freedom ($Z_0 := 1$). When $j = 4$ and $i = 30\,k$ for $k \in \mathbb{N}$, $R_i^{(j)} = |(n^{1/6} Z_i)^{-1}|$ with $(Z_i \mid Z_{i-1} = z)$ following a student distribution with non-centrality parameter z and 3 degrees of freedom ($Z_0 := 1$); else, $R_i^{(j)} = |Z_i| + 1/10$. It is easy to check that when $j = 4$, the assumptions of Theorem 4.4 are not met anymore since R_i^{-1} even lacks finite third order moments when $i = 30\,k$ for some $k \in \mathbb{N}$. Those last two simulation setups can be considered as very simple radial dependence models of the type described in the section 2, where the global volatility at time i depends on the past realizations.

In these 4 settings, we study the empirical null distribution of the multivariate sign test statistic $S^{(n)}(\hat{\boldsymbol{\theta}}^{(n)}_{\text{sign}})$ (see (5)) and of the classical signed-rank test statistic with Van der Waerden scores, which is known to be asymptotically $\chi^2_{p(p+1)/2-1}$ in scenario (I) - see for instance [6]. The Van der Waerden test is a very popular signed-rank test, due to its asymptotic optimality properties under Gaussian assumptions. The use of this test statistic allows us to assess if the robustness properties of $\phi^{(n)}_{\text{sign}}$ are shared by all signed-rank based test statistics. For both test statistics, the location parameter $\boldsymbol{\theta} = \mathbf{0}$ was estimated using $\hat{\boldsymbol{\theta}}^{(n)}_{\text{sign}}$ and the histograms of the empirical null distributions were compared to the target $\chi^2_{p(p+1)/2-1}$ density function. Examination of Figures 1 and 3 confirms the asymptotic validity of $\phi^{(n)}_{\text{sign}}$ and tends to indicate that the Van der Waerden score test probably does not share the same robustness properties. Figures 2 and 4 suggest that the assumptions of Theorems 4.2 and 4.4 are indeed necessary to obtain the asymptotic validity of $\phi^{(n)}_{\text{sign}}$ in scenarios (II) and (III) respectively.

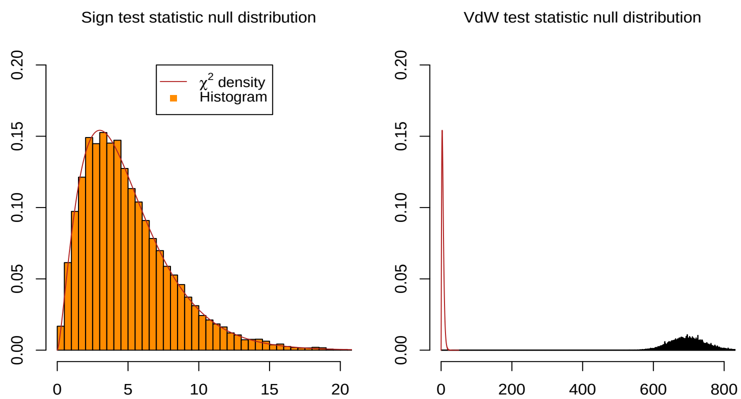

FIG 1. *Histograms of the sign and Van der Waerden test statistics (in orange) under scenario (II) compared to target* χ^2 *density function (in red) when the assumptions of Theorem 4.2 are met. The sample size is* $n = 2,000$.

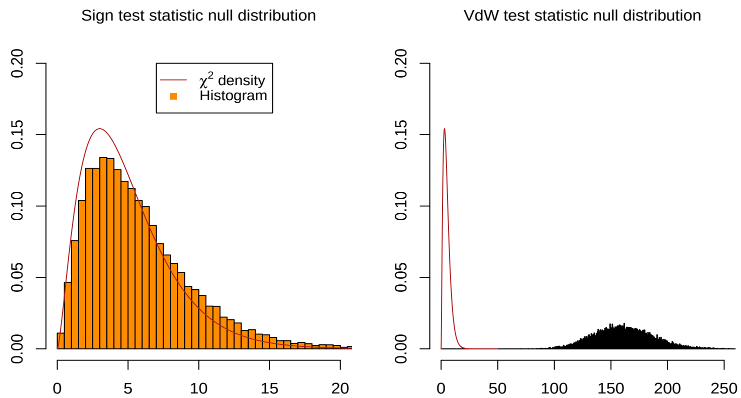

FIG 2. *Histograms of the sign and Van der Waerden test statistics (in orange) under scenario (II) compared to target* χ^2 *density function (in red) when the assumptions of Theorem 4.2 are not met. The sample size is* $n = 2,000$.

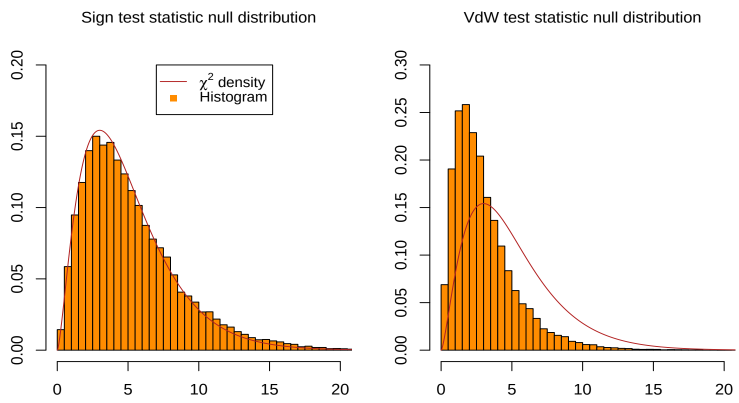

FIG 3. *Histograms of the sign and Van der Waerden test statistics (in orange) under scenario (III) compared to target χ^2 density function (in red) when the assumptions of Theorem 4.4 are met. The sample size is $n = 2,000$.*

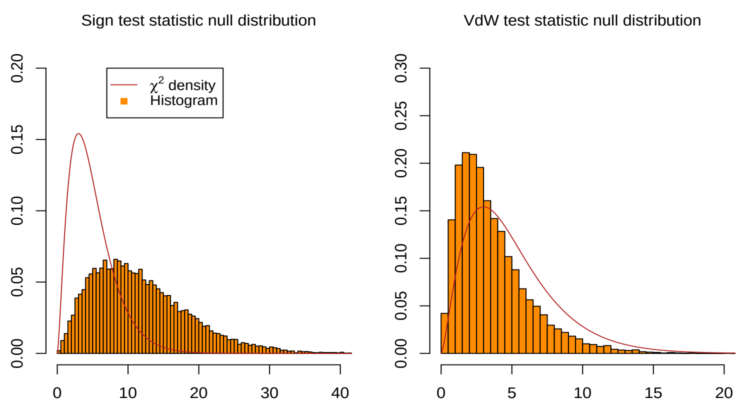

FIG 4. *Histograms of the sign and Van der Waerden test statistics (in orange) under scenario (III) compared to target χ^2 density function (in red) when the assumptions of Theorem 4.4 are not met. The sample size is $n = 2,000$.*

5. Asymptotic power and optimality of the multivariate sign-based test under elliptical directions assumption

5.1. Local asymptotic power in presence of skewness and radial dependence

The asymptotic behavior of $\phi^{(n)}_{\text{sign}}$ under local alternatives is well studied in scenario (I) - see [6]. We then focus on deriving the asymptotic distribution of the test statistic $S^{(n)}(\hat{\boldsymbol{\theta}}^{(n)})$ (see (5)) under local alternatives to $\mathcal{H}_0$ in scenarios (II) and (III). As in previous section, we assume that $\hat{\boldsymbol{\theta}}^{(n)}$ is a root-n consistent estimator of $\boldsymbol{\theta}$.

Theorem 5.1. *Let $\boldsymbol{\tau}^{(n)}$ a converging sequence of $p \times p$ symmetric matrices such that $\det(\mathbf{I}_p + n^{-1/2}\boldsymbol{\tau}^{(n)}) = 1$ and $\mathbf{V}^{(n)} = \mathbf{I}_p + n^{-1/2}\boldsymbol{\tau}^{(n)}$ is positive definite. Let $\boldsymbol{\tau}^{(n)} \to \boldsymbol{\tau}$ as $n \to \infty$. Let $\mathbf{X}_{n1}, \ldots, \mathbf{X}_{nn}$ a triangular array of p-variate elliptical random vector satisfying the assumptions of Theorem 4.2 or 4.4 with shape parameter $\mathbf{V}^{(n)}$ and location parameter $\boldsymbol{\theta}$. This model is denoted $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V}^{(n)},R_1,\ldots,R_n}$.*

Under local alternatives to $\mathcal{H}_0$ of the form $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V}^{(n)},R_1,\ldots,R_n}$, the test statistic $S^{(n)}(\hat{\boldsymbol{\theta}}^{(n)})$ is asymptotically non-central $\chi^2_{p(p+1)/2-1}$ with non-centrality parameter $\frac{p}{2(p+2)}(\mathrm{tr}(\boldsymbol{\tau}^2) - p^{-1}\mathrm{tr}^2(\boldsymbol{\tau}))$ as $n \to \infty$.

Theorem 5.1 immediately yields the local asymptotic power of $\phi^{(n)}_{\text{sign}}$. Interestingly, the asymptotic distribution under scenarios (II) and (III) of $S^{(n)}(\hat{\boldsymbol{\theta}}^{(n)})$ appearing in Theorem 5.1 is the exact same as its asymptotic distribution in scenario (I), derived in [6]. The implication is that when using $\phi^{(n)}_{\text{sign}}$, there is no loss of asymptotic power occasioned by expanding the model from the classical elliptical setting (I) to scenarios involving skewness and radial dependence considered in this contribution. We should stress out that this result is again rather powerful since the R_i could belong to some very general class of stochastic processes.

5.2. Asymptotic optimality

We will now turn to the question of the optimality of the multivariate sign test $\phi^{(n)}_{\text{sign}}$. Following Le Cam asymptotic theory of experiments, the main tool will be a Local Asymptotic Normality (LAN) result for the sign-based central sequence (4). In the classical elliptical framework (I), it is common to consider the following slightly more stringent *elliptical density* model where the R_i are such that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ admit a density with respect to the Lebesgue measure. In this model, denoted by $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_1}$, $\mathbf{X}_1, \ldots, \mathbf{X}_n$, have common density

$$f_{\mathbf{X}}(\mathbf{x}) = c_{p,\sigma,f_1} f_1\left(\frac{1}{\sigma}\big((\mathbf{x}-\boldsymbol{\theta})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\theta})\big)^{1/2}\right),$$

for some *scale parameter* $\sigma \in \mathbb{R}_0^+$, some $f_1 \in W^{1,2}(\mathbb{R}_0^+)$ belonging to the class of the *standardized radial functions*

$$\mathcal{F}_1 := \{f : \mathbb{R}_0^+ \to \mathbb{R}^+ \mid (\int_0^\infty r^{p-1} f(r)dr < \infty) \cap (\frac{\int_0^1 r^{p-1} f(r)dr}{\int_0^\infty r^{p-1} f(r)dr} = 1/2)\}$$

and c_{p,σ,f_1} a normalization constant. This purely technical constraint on f_1 allows to avoid identifiability issues without any moment hypothesis. In this context, it is well known that there is no radial function belonging to $\mathcal{F}_1$ such that the sign-based test for sphericity $\phi^{(n)}_{\text{sign}}$ is locally and asymptotic optimal. However, given a radial function $f_1 \in \mathcal{F}_1$, the following LAN property holds, as shown by [6]. This LAN property allows us to construct f_1-specified locally and asymptotically optimal tests for $\mathcal{H}_0$ based on the *central sequence* $\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\theta}, \sigma, \mathbf{V})$ defined in the Theorem 5.2 below.

Theorem 5.2. *For any $f_1 \in \mathcal{F}_1$ and for any sequence $\boldsymbol{\tau}^{(n)}$ as in Theorem 5.1 we have under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_1}$ and as $n \to \infty$ that*

$$\begin{aligned}\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}+n^{-1/2}\boldsymbol{\tau}^{(n)},f_1}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_1}}\right) &= \mathrm{ve\mathring{c}h}\,'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\theta},\sigma,\mathbf{V}) \\ &\quad - \frac{1}{2}\mathrm{ve\mathring{c}h}\,'(\boldsymbol{\tau})\boldsymbol{\Gamma}_{f_1}(\mathbf{V})\mathrm{ve\mathring{c}h}\,(\boldsymbol{\tau}) + o_{\mathrm{P}}(1),\end{aligned}$$

where

$$\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\theta},\sigma,\mathbf{V}) = \frac{n^{1/2}}{2}\mathbf{M}_p(\mathbf{V})\mathrm{vec}(\mathbf{V}^{-1/2}\mathbf{S}^{(n)}_{f_1}(\boldsymbol{\theta},\sigma,\mathbf{V})\mathbf{V}^{-1/2}),$$

$$\mathbf{S}^{(n)}_{f_1}(\boldsymbol{\theta},\sigma,\mathbf{V}) = n^{-1}\sum_{i=1}^n -\frac{\dot{f}_1}{f_1}\big(\frac{d_i(\boldsymbol{\theta},\mathbf{V})}{\sigma}\big)\frac{d_i(\boldsymbol{\theta},\mathbf{V})}{\sigma}\mathbf{U}_i(\boldsymbol{\theta},\mathbf{V})\mathbf{U}_i(\boldsymbol{\theta},\mathbf{V})',$$

$$\begin{aligned}\boldsymbol{\Gamma}_{f_1}(\mathbf{V}) &= \frac{J_p(f_1)}{4p(p+2)}\mathbf{M}_p(\mathbf{V})\,(\mathbf{V}^{\otimes 2})^{-1/2}\big(\mathbf{I}_{p^2} + \mathbf{K}_p - \frac{2}{p}\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}(\mathbf{I}_p)'\big) \\ &\quad (\mathbf{V}^{\otimes 2})^{-1/2}\,\mathbf{M}_p'(\mathbf{V})\end{aligned}$$

and

$$J_p(f_1) = \frac{\int_0^\infty \frac{\dot{f}_1^2(r)}{f_1(r)} r^{p+1}dr}{\int_0^\infty f_1(r)r^{p-1}dr}.$$

Moreover, $\boldsymbol{\Delta}^{(n)}_{f_1}(\boldsymbol{\theta},\sigma,\mathbf{V}) \to \mathcal{N}(\mathbf{0},\boldsymbol{\Gamma}_{f_1}(\mathbf{V}))$ as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_1}$.

As mentioned earlier, this theorem does not allow us to derive an optimality property for ϕ_{sign} since there is no radial function $f_1 \in \mathcal{F}_1$ satisfying $\mathbf{\Delta}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{I}_p) = \mathbf{\Delta}^{(n)}_{f_1}(\boldsymbol{\theta}, \sigma, \mathbf{I}_p)$. We then take any positive real sequence $0 < \nu_n = o(1)$ and consider the sequence of radial functions

$$f_n(r) = (1 + \frac{r^2}{\nu_n})^{-\frac{p+\nu_n}{2}} \tag{8}$$

with

$$c_{p,\sigma,f_n} = \frac{\Gamma((p+\nu_n)/2)}{\pi^{p/2}\nu_n^{p/2}\sigma^p\,\Gamma(\nu_n/2)}.$$

These radial functions correspond to random vectors following multivariate t-distributions with $\nu_n > 0$ degrees of freedom. Even if f_n does not technically belong to $\mathcal{F}_1$, there is no identifiability problem in this specific scenario. Later, when we refer to (8), the sequence can in some situations be standardized such that it belongs to $\mathcal{F}_1$. We have an equivalent of the classical LAN property of Theorem 5.2 for this sequence f_n of radial functions.

Theorem 5.3. *Let $\mathbf{V} = \mathbf{I}_p$. Let $\mathbf{X}_{n1}, \ldots, \mathbf{X}_{nn}$ a triangular array of i.i.d. random vectors with distribution $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$. For any sequence of radial functions f_n satisfying (8) and for any sequence $\boldsymbol{\tau}^{(n)}$ as in Theorem 5.1, we have under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$ and as $n \to \infty$ that*

$$\begin{aligned}\log\left(\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}+n^{-1/2}\boldsymbol{\tau}^{(n)},f_n}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) &= \overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\mathbf{\Delta}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{V}) \\ &\quad - \frac{1}{2}\overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\mathbf{\Gamma}_{\text{sign}}(\mathbf{V})\overset{\circ}{\mathrm{vech}}(\boldsymbol{\tau}) + o_{\mathrm{P}}(1),\end{aligned}$$

where

$$\mathbf{\Delta}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{V}) := \frac{n^{1/2}}{2}\mathbf{M}_p(\mathbf{V})\mathrm{vec}(\mathbf{V}^{-1/2}\mathbf{S}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{V})\mathbf{V}^{-1/2}),$$

and

$$\begin{aligned}\mathbf{\Gamma}_{\text{sign}}(\mathbf{V}) &:= \frac{p}{4(p+2)}\mathbf{M}_p(\mathbf{V})\,(\mathbf{V}^{\otimes 2})^{-1/2}\big(\mathbf{I}_{p^2} + \mathbf{K}_p - \frac{2}{p}\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}(\mathbf{I}_p)'\big) \\ &\quad (\mathbf{V}^{\otimes 2})^{-1/2}\,\mathbf{M}'_p(\mathbf{V}).\end{aligned}$$

Moreover, $\mathbf{\Delta}^{(n)}_{\text{sign}}(\boldsymbol{\theta}, \mathbf{V}) \to \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_{\text{sign}}(\mathbf{V}))$ as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$.

A proof is given in the Appendix. Note that the rate of convergence of ν_n plays no role in Theorem 5.3, which is rather uncommon in this type of asymptotic results. Now, Theorem 5.3 alongside classical Le Cam asymptotic theory of experiments directly yields the following asymptotic optimality result for the multivariate sign test for sphericity $\phi^{(n)}_{\text{sign}}$.

Theorem 5.4. *For any sequence of radial functions f_n satisfying (8) and under the sequence of experiments $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$ as in Theorem 5.3, the test $\phi^{(n)}_{\mathrm{sign}}$ is locally and asymptotically maximin when testing for $\mathcal{H}_0$ against alternatives of the type $\mathbf{V}^{(n)} = \mathbf{I}_p + n^{-1/2}\boldsymbol{\tau}^{(n)}$, with $\boldsymbol{\tau}^{(n)}$ as in Theorem 5.1.*

The sequence of experiments corresponding to radial functions (8) is such that, as $n \to \infty$, the random vectors considered tend to have increasingly heavy tails. An intuitive way to view the optimality result of Theorem 5.4 is that the heavier the tails of the considered multivariate student random vectors, the better the asymptotic performance of the sign test compared to its competitors. We emphasize the fact that, apart from the trivially degenerate case $\nu_n = 0$, there is no rate of convergence such that $\phi^{(n)}_{\mathrm{sign}}$ is not asymptotically optimal. As mentioned in the introduction, this result is rather interesting in itself since getting rid of all the radial information present in the random vectors turns out to be the optimal choice in this scenario. This fact is rather counter-intuitive in the sense that getting rid of a large part of the - seemingly relevant - information is the asymptotically optimal way to perform inference in this context. As the assumptions of Theorem 5.4 may seem a bit too restrictive, we introduce a more general result, valid for any collection of random vectors with elliptical directions such that its limiting behavior is similar to that of Theorem 5.4. We consider the following triangular array of random vectors with elliptical directions, generalizing the model used in proposition 5.4.

Let $\mathbf{X}_{n1}, \ldots, \mathbf{X}_{nn}$ a triangular array of i.i.d. random p-vectors with elliptical directions (see (1)) such that, for every n, the array admits joint density

$$
\begin{aligned}
f_{\mathbf{X}_{n1},\ldots,\mathbf{X}_{nn}}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \quad = \quad & c_{p,\sigma,\tilde{f}_n} \tilde{f}_n \bigg(\frac{1}{\sigma} \big((\mathbf{x}_1 - \boldsymbol{\theta})' \mathbf{V}^{-1} (\mathbf{x}_1 - \boldsymbol{\theta}) \big)^{1/2}, \ldots, \\
& \frac{1}{\sigma} \big((\mathbf{x}_n - \boldsymbol{\theta})' \mathbf{V}^{-1} (\mathbf{x}_n - \boldsymbol{\theta}) \big)^{1/2} \bigg).
\end{aligned}
$$

The sequence of *joint radial functions* are assumed to belong to the class

$$
\begin{aligned}
\tilde{\mathcal{F}}_1 \quad = \quad & \{ \tilde{f}_1 : (\mathbb{R}_0^+)^p \to \mathbb{R}^+ \mid (\int_{(\mathbb{R}_0^+)^p} \prod_{i=1}^n r_i^{p-1} \tilde{f}_1(r_1, \ldots, r_n) dr_1 \ldots dr_n < \infty) \\
& \cap (\frac{\int_0^1 \cdots \int_0^1 \prod_{i=1}^n r_i^{p-1} \tilde{f}_1(r_1, \ldots, r_n) dr_1 \ldots dr_n}{\int_0^\infty \cdots \int_0^\infty \prod_{i=1}^n r_i^{p-1} \tilde{f}_1(r_1, \ldots, r_n) dr_1 \ldots dr_n} = 1/2) \}. \qquad (9)
\end{aligned}
$$

This model is denoted $\tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}$ and includes some cases from both scenarios (I) and (III). It can be viewed as a generalized version of the classical elliptical density model where some dependence may be present in the radial part of the random vectors.

Theorem 5.5. *Let $\tilde{f}_n$ a sequence of joint radial functions belonging to the class (9). Assume that there exist a sequence $0 < \nu_n = o(1)$ with corresponding (standardized) sequence f_n satisfying (8) such that*

$$\left\| \frac{\tilde{f}_n(r_1, \ldots, r_n)}{\prod_{i=1}^n f_n(r_i)} \right\|_\infty = 1 + o(1)$$

as $n \to \infty$.

Then, under the sequence of experiments $\tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}$*, the test* $\phi^{(n)}_{\mathrm{sign}}$ *is locally and asymptotically maximin when testing for* $\mathcal{H}_0$ *against alternatives of the type* $\mathbf{V}^{(n)} = \mathbf{I}_p + n^{-1/2}\boldsymbol{\tau}^{(n)}$*, with* $\boldsymbol{\tau}^{(n)}$ *as in Theorem 5.1.*

Theorem 5.5 is a slightly more general version of Theorem 5.4 which applies to arrays of - potentially dependent - elliptical random vectors such that their joint radial densities are sufficiently close to the joint radial densities of independent multivariate students random vectors with increasingly heavy tails. Note that there is no need to require additional conditions in terms of the speed of convergence of $\tilde{f}_n(r_1, \ldots, r_n)/\prod_{i=1}^n f_n(r_i)$ to obtain local asymptotic optimality of $\phi^{(n)}_{\mathrm{sign}}$. In this sense, Theorem 5.5 is a fairly general result, valid for any collection of random vectors which tend to behave asymptotically like multivariate student random vectors with a very small number of degrees of freedom. The main point of Theorem 5.5 is to establish that the asymptotic optimality of $\phi^{(n)}_{\mathrm{sign}}$ does not depend on the speed of convergence of $\tilde{f}_n$, even in a class of radial functions slightly more general than $\mathcal{F}_1$.

5.3. Simulation study

We now present some simulation results illustrating the fact that the local asymptotic power of $\phi^{(n)}_{\mathrm{sign}}$ are the same in scenarios (I), (II) and (III).

We generated $M = 1,000$ independent samples $\mathbf{X}_1^{(j,\tau)}, \ldots, \mathbf{X}_n^{(j,\tau)}$ with $n = 10,000$, $\tau \in \{0, 1, \ldots, 9, 10\}$ and $j \in \{1, 2, 3\}$. For all j, i and τ, the $(p = 3)$-dimensional random vector satisfies $\mathbf{X}_i^{(j)} = R_i^{(j)}(\frac{\mathbf{V}_\tau^{(n)}}{\det(\mathbf{V}_\tau^{(n)})^{1/p}})^{1/2}\mathbf{U}_i$ with

$$\mathbf{V}_\tau^{(n)} = \mathrm{diag}(1, 1, 1 - n^{-1/2}\tau)$$

and with the $\mathbf{U}_i$ uniformly distributed on $\mathcal{S}^2$. For $\tau = 0$, the data-generating process is spherical while for every $\tau > 0$, the data-generating process is increasingly under the alternative.

The case $j = 1$ corresponds to scenario (I): $R_i^{(j)} = |Z_i|$ with Z_i i.i.d. standard Gaussian random variables with expectation 1. The case $j = 2$ corresponds to scenario (II): $R_i^{(j)} = |Z_i| + |\langle (0, 0, 1), \mathbf{U}_i' \rangle| + 1/10$ where Z_i are i.i.d. standard Gaussian random variables with expectation 1. Finally, the case $j = 3$ corresponds to scenario (III): $R_i^{(j)} = |Z_i| + 1/10$ with $(Z_i \mid Z_{i-1} = z) \sim \mathcal{N}(z, 1)$ and $Z_0 := 1$.

In the three settings, we study the empirical rejection frequencies of the multivariate sign test $\phi^{(n)}_{\mathrm{sign}}$, compared to the theoretical asymptotic local power derived in Theorem 5.1. The test $\phi^{(n)}_{\rm sign}$ is performed at the asymptotic confidence level $\alpha = 0.05$. Inspection of Figure 5 confirms the theoretical findings of Theorem 5.1.

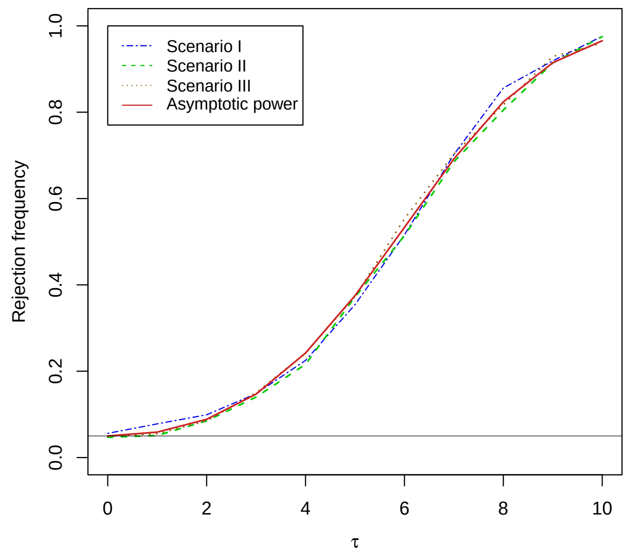

Fig 5. *Empirical rejection frequencies of $\phi^{(n)}_{\rm sign}$ under scenario (I) (in dotted blue), scenario (II) (in dotted green) and scenario (III) (in dotted orange) compared to the theoretical asymptotic power curve (in red). The sample size is $n = 10,000$.*

We conclude this section by a last simulation study meant to illustrate the asymptotic optimality properties of $\phi^{(n)}_{\rm sign}$. We generated $M = 2,000$ independent samples $\mathbf{X}^{(\tau)}_1, \ldots, \mathbf{X}^{(\tau)}_n$ with $n \in \{110, 200, 500, 1000\}$ and $\tau \in \{0, 1, \ldots, 9, 10\}$. For all τ and for all n, $\mathbf{X}^{(\tau)}_1, \ldots, \mathbf{X}^{(\tau)}_n$ are i.i.d. and are drawn according to a $(p = 3)$-dimensional multivariate t-distribution with $\frac{150}{n}$ degrees of freedom and shape parameter $\frac{\mathbf{V}^{(n)}_\tau}{\det(\mathbf{V}^{(n)}_\tau)^{1/p}}$ with $\mathbf{V}^{(n)}_\tau = \mathrm{diag}(1, 1, 1 - n^{-1/2}\tau)$. The larger n is, the larger the dispersion of the data and the closest the simulation setup is to the model of Theorem 5.4, where $\phi^{(n)}_{\rm sign}$ is locally and asymptotically optimal. We performed at the asymptotic level α the multivariate sign test $\phi^{(n)}_{\rm sign}$ and the signed-rank test with Van der Waerden score. Inspection of Figure 6 confirms that $\phi^{(n)}_{\rm sign}$ tends to outperform more and more the Van der Waerden test as the tails of the distribution grow heavier (with n) - which is in line with Theorem 5.4.

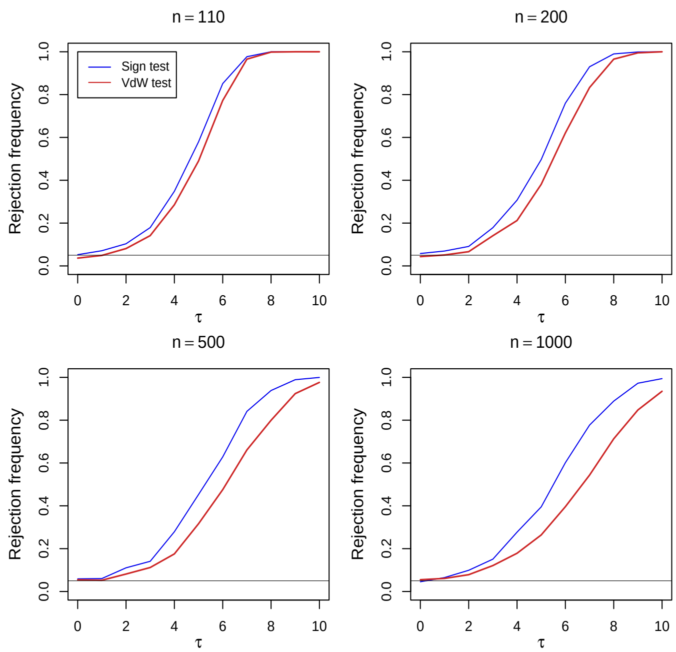

FIG 6. *Empirical rejection frequencies of* $\phi_{\mathrm{sign}}^{(n)}$ *(in blue) and of the Van der Waerden test (in red) as a function of* τ *under multivariate student assumption with df* $= 150/n$ *(scenario (I)). The sample size is* $n \in \{110, 200, 500, 1000\}$.

6. Conclusions

In this contribution, we have studied the asymptotic behavior of the spatial sign test for sphericity $\phi_{\text{sign}}^{(n)}$ when the model does not satisfy the classical elliptical distribution assumption. Expanding on the elliptical assumption, we have identified two settings in which testing for sphericity is a highly relevant problem that can be reframed through a parametric lens: elliptical directions in presence of skewness (II) and elliptical directions with radial dependence (III). We have derived asymptotic validity results and local asymptotic power results for $\phi_{\text{sign}}^{(n)}$ in these two settings. In particular, we have shown that a very general type of dependence can be present in the radial part of the observations without affecting the asymptotic validity and local power of $\phi_{\text{sign}}^{(n)}$. We have also shown that $\phi_{\text{sign}}^{(n)}$ enjoys certain asymptotic optimality properties when the underlying process tends to be very strongly heavy tailed. These results imply that, assuming a strongly heavy tailed data-generating process, the use of the spatial sign test for sphericity will guarantee exceptional asymptotic power while providing some asymptotic robustness to skewness and radial dependence. It should be noted that these robustness properties do not seem to be enjoyed by every test in the class of signed-rank tests proposed by [6] - as highlighted in the simulation study in section 4. This is why the combination of Theorems 4.2, 4.4 and 5.4 argues strongly in favour of using $\phi_{\text{sign}}^{(n)}$ when the data generating process is suspected to be strongly heavy-tailed. Indeed, from a practical point of view, the asymptotic power of $\phi_{\text{sign}}^{(n)}$ will be arbitrarily close to the asymptotic power of the best signed-rank based competitors, with stronger guarantees regarding its robustness to the lack of classical elliptical assumption. When n is finite, the asymptotic optimality of the test $\phi_{\text{new}}^{(n)}$ results in a tendency for $\phi_{\text{sign}}^{(n)}$ to dominate in terms of power the classical signed-rank based competitors when the data generating process is strongly heavy-tailed. This fact was highlighted in the simulation study in the section 5. In this sense, the asymptotic results of this contribution are of practical use as they admit a fairly straightforward interpretation for finite n. An interesting open question concerns the conditions under which procedures based on the signed-rank - such as the Van der Waerden test considered in the various simulation studies - would exhibit the same type of robustness as $\phi_{\text{sign}}^{(n)}$ in scenarios (II) and (III). The simulation study of the section 4 suggests that these conditions would be different from those derived for $\phi_{\text{sign}}^{(n)}$ in this contribution. This question is left to future research.

Appendix: Proofs of the various results

Proof of Theorem 4.1. The proof consists in a direct application of Theorem 2 in [3]. □

Proof of Theorem 4.2. The result is a direct consequence of Theorem 4.1. □

Proof of Theorem 4.3. First note that if $n^{-3/2}\sum_{i=1}^{n} R_i^{-3} \xrightarrow{\mathrm{P}} 0$, then we have by Jensen inequality and the fact that $R_i \geq 0$ almost surely that

$$0 \leq (n^{-3/2}\sum_{i=1}^{n} R_i^{-2})^{3/2} \leq n^{-3/4}(n^{-1}\sum_{i=1}^{n} R_i^{-2})^{3/2} \leq n^{-3/4}n^{-1}\sum_{i=1}^{n} R_i^{-3} \xrightarrow{\mathrm{P}} 0.$$

This directly implies that $n^{-3/2}\sum_{i=1}^{n} R_i^{-2} = o_{\mathrm{P}}(1)$, a fact that will be used later. Now, we note that

$$\begin{aligned}
&\frac{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2 - 2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}) + \|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\mathbf{U}_i'(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\\
= \;& \frac{\|\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\mathbf{U}_i'(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\\
= \;& \frac{\|\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\frac{(\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)})(\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)})'}{\|\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)}\|^2}\\
= \;& \frac{(\mathbf{X}_i-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2} + \frac{(\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)})(\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)})' - (\mathbf{X}_i-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\\
= \;& \mathbf{U}_i(\boldsymbol{\theta},\mathbf{I}_p)\mathbf{U}_i'(\boldsymbol{\theta},\mathbf{I}_p) + \frac{(\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)})(\mathbf{X}_i-\hat{\boldsymbol{\theta}}^{(n)})' - (\mathbf{X}_i-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\\
= \;& \mathbf{U}_i(\boldsymbol{\theta},\mathbf{I}_p)\mathbf{U}_i'(\boldsymbol{\theta},\mathbf{I}_p)\\
&+\frac{(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' - (\mathbf{X}_i-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' - (\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2},
\end{aligned}$$

which yields

$$\begin{aligned}
&\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\mathbf{U}_i'(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\\
= \;& \mathbf{U}_i(\boldsymbol{\theta},\mathbf{I}_p)\mathbf{U}_i'(\boldsymbol{\theta},\mathbf{I}_p)\\
&+\frac{(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' - (\mathbf{X}_i-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' - (\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\\
&+\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}) - \|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\mathbf{U}_i'(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p).
\end{aligned} \tag{10}$$

Now, (10) implies that

$$\begin{aligned}
n^{1/2}(\mathbf{S}^{(n)}_{\text{sign}}(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)-\mathbf{S}^{(n)}_{\text{sign}}(\boldsymbol{\theta},\mathbf{I}_p)) &= n^{-1/2}\sum_{i=1}^{n}\Big((\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' \\
&\quad -(\mathbf{X}_i-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' \\
&\quad -(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'\Big)\frac{1}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2} \\
&\quad +n^{-1/2}\sum_{i=1}^{n}\left(\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\right. \\
&\quad \left.-\frac{\|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\right)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i'(\hat{\boldsymbol{\theta}},\mathbf{I}_p) \\
&:= \mathbf{A}_n+\mathbf{B}_n.
\end{aligned}$$

Recall that it has been established at the very beginning of the proof that as $n\to\infty$, $n^{-3/2}\sum_{i=1}^{n}R_i^{-2}=o_{\mathrm{P}}(1)$. Now, using this last fact and $n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})=O_{\mathrm{P}}(1)$, we get that as $n\to\infty$,

$$\begin{aligned}
\mathbf{A}_n &= n^{-1/2}\sum_{i=1}^{n}\|\mathbf{X}_i-\boldsymbol{\theta}\|^{-2}\Big((\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})'-(\mathbf{X}_i-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' \\
&\quad -(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'\Big) \\
&= n^{-1/2}\sum_{i=1}^{n}R_i^{-1}\|\mathbf{V}^{1/2}\mathbf{U}_i\|^{-2}\Big(R_i^{-1}(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' \\
&\quad -\mathbf{V}^{1/2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})'-(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})\mathbf{U}_i'\mathbf{V}^{1/2}\Big) \\
&= -n^{-1/2}\sum_{i=1}^{n}R_i^{-1}\|\mathbf{V}^{1/2}\mathbf{U}_i\|^{-2}\Big(\mathbf{V}^{1/2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})'+(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})\mathbf{U}_i'\mathbf{V}^{1/2}\Big) \\
&\quad +o_{\mathrm{P}}(1).
\end{aligned} \tag{11}$$

Using the independence of the R_i and the $\mathbf{U}_i$ (which are i.i.d.) alongside $n^{-3/2}\sum_{i=1}^{n}R_i^{-2}=o_{\mathrm{P}}(1)$, we get that as $n\to\infty$

$$\begin{aligned}
\mathrm{E}\big(\|n^{-1}\sum_{i=1}^{n} R_i^{-1}\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\|^2\big) &= \mathrm{E}\big(n^{-2}\sum_{i=1}^{n} R_i^{-1}\frac{\mathbf{U}_i'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2} \\
&\quad \sum_{j=1}^{n} R_j^{-1}\frac{\mathbf{V}^{1/2}\mathbf{U}_j}{\|\mathbf{V}^{1/2}\mathbf{U}_j\|^2}\big) \\
&= n^{-2}\sum_{i=1}^{n}\mathrm{E}(R_i^{-2})\mathrm{E}\big(\frac{\mathbf{U}_i'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\big) \\
&\quad + n^{-2}\sum_{1\le i\ne j\le n}\Big(\mathrm{E}(R_i^{-1}R_j^{-1}) \\
&\quad \mathrm{E}\big(\frac{\mathbf{U}_i'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\frac{\mathbf{V}^{1/2}\mathbf{U}_j}{\|\mathbf{V}^{1/2}\mathbf{U}_j\|^2}\big)\Big) \\
&= \mathrm{E}\big(\frac{\mathbf{U}_1'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_1\|^2}\frac{\mathbf{V}^{1/2}\mathbf{U}_1}{\|\mathbf{V}^{1/2}\mathbf{U}_1\|^2}\big)\; n^{-2}\sum_{i=1}^{n}\mathrm{E}(\frac{1}{R_i^2}) \\
&= o(1). \qquad (12)
\end{aligned}$$

It is implied by (12) that $n^{-1}\sum_{i=1}^{n} R_i^{-1}\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2} = o_{\mathrm{P}}(1)$. Then, using again that $n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}) = O_{\mathrm{P}}(1)$, we get from this last fact and from (11) that

$$\begin{aligned}
\mathbf{A}_n &= -n^{-1/2}\sum_{i=1}^{n}\frac{\mathbf{V}^{1/2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' + (\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})\mathbf{U}_i'\mathbf{V}^{1/2}}{R_i\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2} + o_{\mathrm{P}}(1) \\
&= -n^{-1}\sum_{i=1}^{n}\frac{\mathbf{V}^{1/2}\mathbf{U}_i\; n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})' + n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})\mathbf{U}_i'\mathbf{V}^{1/2}}{R_i\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2} + o_{\mathrm{P}}(1) \\
&= o_{\mathrm{P}}(1)
\end{aligned}$$

as $n \to \infty$. Now, consider

$$\mathbf{B}_n = n^{-1/2}\sum_{i=1}^{n}\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}) - \|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)'.$$

Using arguments of the same type as earlier and noting that

$$\begin{aligned}
\Big\|n\|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2 n^{-3/2}\sum_{i=1}^{n}\frac{1}{R_i^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)'\Big\|_{\mathrm{F}} &\le n\|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2 \\
&\quad \Big\|n^{-3/2}\sum_{i=1}^{n}\frac{1}{R_i^2}\mathbf{1}_p\mathbf{1}_p'\Big\|_{\mathrm{F}} \\
&= o_{\mathrm{P}}(1),
\end{aligned}$$

we get that

$$\begin{aligned}
\mathbf{B}_n &= n^{-1/2}\sum_{i=1}^{n}\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)' \\
&\quad -n^{-1/2}\sum_{i=1}^{n}\frac{\|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)' \\
&= n^{-1/2}\sum_{i=1}^{n}\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)' \\
&\quad -n\|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2 n^{-3/2}\sum_{i=1}^{n}\frac{1}{R_i^2\lambda_{\mathbf{V},p}}\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)' \\
&= n^{-1/2}\sum_{i=1}^{n}\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)\mathbf{U}_i(\hat{\boldsymbol{\theta}},\mathbf{I}_p)'+o_{\mathrm{P}}(1),
\end{aligned}$$

with $\lambda_{\mathbf{V},p}$ the last eigenvalue of $\mathbf{V}$. We then have, using (10) again, that as $n\to\infty$

$$\begin{aligned}
\mathbf{B}_n &= n^{-1/2}\sum_{i=1}^{n}\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\big(\mathbf{U}_i(\boldsymbol{\theta},\mathbf{I}_p)\mathbf{U}_i'(\boldsymbol{\theta},\mathbf{I}_p) \\
&\quad +\frac{(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})'-(\mathbf{X}_i-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})'-(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2} \\
&\quad +\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})-\|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\mathbf{U}_i'(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\big)+o_{\mathrm{P}}(1).
\end{aligned} \tag{13}$$

We denote by $\tilde{\mathbf{U}}_i^{(\ell)}$ the ℓ-th component of $\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|}$, $1\leq\ell\leq p$. Now, note that for the first term of (13), we have for $1\leq k,l\leq p$ that

$$\begin{aligned}
&(n^{-1/2}\sum_{i=1}^{n}\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\boldsymbol{\theta},\mathbf{I}_p)\mathbf{U}_i'(\boldsymbol{\theta},\mathbf{I}_p))_{k,l} \\
&= 2\sum_{m=1}^{p}n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})^{(m)}n^{-1}\sum_{i=1}^{n}R_i^{-1}\lambda_{\mathbf{V},p}^{-1/2}\tilde{\mathbf{U}}_i^{(m)}\tilde{\mathbf{U}}_i^{(k)}\tilde{\mathbf{U}}_i^{(l)}.
\end{aligned}$$

Now, using again that $n^{-3/2}\sum_{i=1}^{n}R_i^{-2}=o_{\mathrm{P}}(1)$, we have that

$$n^{-1}\sum_{i=1}^{n}R_i^{-1}\tilde{\mathbf{U}}_i^{(m)}\tilde{\mathbf{U}}_i^{(k)}\tilde{\mathbf{U}}_i^{(l)}$$

converges to 0 in quadratic mean because as $n \to \infty$,

$$\begin{aligned}
\mathrm{E}\big((n^{-1}\sum_{i=1}^{n} R_i^{-1}\tilde{\mathbf{U}}_i^{(m)}\tilde{\mathbf{U}}_i^{(k)}\tilde{\mathbf{U}}_i^{(l)})^2\big) &= \mathrm{E}(n^{-2}\sum_{i=1}^{n} R_i^{-1}\tilde{\mathbf{U}}_i^{(m)}\tilde{\mathbf{U}}_i^{(k)}\tilde{\mathbf{U}}_i^{(l)} \\
&\quad \sum_{i=1}^{n} R_j^{-1}\tilde{\mathbf{U}}_j^{(m)}\tilde{\mathbf{U}}_j^{(k)}\tilde{\mathbf{U}}_j^{(l)}) \\
&\leq n^{-2}\sum_{1\leq i\neq j\leq n}\Big(\mathrm{E}(R_i^{-1}R_j^{-1}) \\
&\quad \mathrm{E}(\tilde{\mathbf{U}}_i^{(m)}\tilde{\mathbf{U}}_i^{(k)}\tilde{\mathbf{U}}_i^{(l)}\tilde{\mathbf{U}}_j^{(m)}\tilde{\mathbf{U}}_j^{(k)}\tilde{\mathbf{U}}_j^{(l)})\Big) \\
&\quad +n^{-2}\sum_{i=1}^{n}\mathrm{E}(R_i^{-2}) \\
&= o(1).
\end{aligned} \tag{14}$$

The last line holds because $\mathrm{E}(\tilde{\mathbf{U}}_i^{(m)}\tilde{\mathbf{U}}_i^{(l)}\tilde{\mathbf{U}}_i^{(k)}) = -\mathrm{E}(\tilde{\mathbf{U}}_i^{(m)}\tilde{\mathbf{U}}_i^{(l)}\tilde{\mathbf{U}}_i^{(k)}) = 0$ (recall that the $\tilde{\mathbf{U}}_i$ are i.i.d.). Then, (13) and (14) yield

$$\begin{aligned}
\mathbf{B}_n &= n^{-1/2}\sum_{i=1}^{n}\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2} \\
&\quad \Big(\frac{(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})'-(\mathbf{X}_i-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})'-(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})(\mathbf{X}_i-\boldsymbol{\theta})'}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2} \\
&\quad +\frac{2(\mathbf{X}_i-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta})-\|\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\|^2}{\|\mathbf{X}_i-\boldsymbol{\theta}\|^2}\mathbf{U}_i(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\mathbf{U}_i'(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)\Big)+o_{\mathrm{P}}(1).
\end{aligned}$$

The fact that $n^{-3/2}\sum_{i=1}^{n} R_i^{-3} = o_{\mathrm{P}}(1)$, $n^{-3/2}\sum_{i=1}^{n} R_i^{-2} = o_{\mathrm{P}}(1)$ and that $n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}) = O_{\mathrm{P}}(1)$ alongside same reasoning as earlier yields $\mathbf{B}_n = o_{\mathrm{P}}(1)$. Then, as $n \to \infty$, we have that

$$n^{1/2}(\mathbf{S}_{\mathrm{sign}}^{(n)}(\hat{\boldsymbol{\theta}}^{(n)},\mathbf{I}_p)-\mathbf{S}_{\mathrm{sign}}^{(n)}(\boldsymbol{\theta},\mathbf{I}_p)) = \mathbf{A}_n+\mathbf{B}_n = o_{\mathrm{P}}(1),$$

which concludes the proof. □

Proof of Theorem 4.4. The result is a direct consequence of Theorem 4.3. □

Proof of Theorem 4.5. The proof holds trivially because, by definition of c_n,

$$\begin{aligned}
\lim_{n\to\infty} n^{-3/2}\sum_{i=1}^{n} R_{in}^{-3} &\leq \lim_{n\to\infty} \max_{i\in\{1,\ldots,n\}}\{n^{-1/2}R_{in}^{-3}\} \\
&= \lim_{n\to\infty} \max_{i\in\{1,\ldots,n\}}\{n^{-1/2}(|A_i|+c_n)^{-3}\} \\
&\leq \lim_{n\to\infty} n^{-1/2}c_n^{-3} \\
&= \lim_{n\to\infty} (n^{-1/6}c_n^{-1})^3 \\
&= 0.
\end{aligned}$$

Note that it is clear that the result hold for any stochastic process A_i.

□

Proof of Theorem 4.6. Note that since $p \geq 7$ and because the $\mathbf{X}_i$ have common (bounded) density $f_{\mathbf{X}_1}$, $\mathrm{E}(R_i^{-6}) = \int_0^\infty r^{-6}r^{p-1}f_{\mathbf{X}_1}(r)\mathrm{d}r < \infty$. Now, write

$$\lim_{n\to\infty} n^{-3/2}\sum_{i=1}^{n} R_i^{-3} \leq \lim_{n\to\infty} \max_{i\in\{1,\ldots,n\}}\{n^{-1/2}R_i^{-3}\}.$$

We also note that $\max_{i\in\{1,\ldots,n\}}\{n^{-1/2}R_i^{-3}\} = o_{\mathrm{P}}(1)$ if

$$\max_{i\in\{1,\ldots,n\}}\{n^{-1}R_i^{-6}\} = o_{\mathrm{P}}(1)$$

as $n \to \infty$. Now, since the R_i are identically distributed, we have that for any fixed $\delta \in \mathbb{R}_0^+$,

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{P}(R_i^{-6} \geq i\delta) = \lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{P}(R_1^{-6} \geq i\delta) \leq \frac{\mathrm{E}(R_1^{-6})}{\delta} < \infty.$$

Then, by the Borel-Cantelli lemma, we have that for any arbitrary $\delta > 0$,

$$\mathrm{P}([n^{-1}R_n^{-6} \geq \delta] \text{ infinitely often}) = 0.$$

In other words, $n^{-1}R_n^{-6}$ converges almost surely to 0 as $n \to \infty$. It remains to check that as $n \to \infty$, $\max_{i\in\{1,\ldots,n\}}\{n^{-1}R_i^{-6}\} = o_{\mathrm{P}}(1)$ (actually, we show the almost sure convergence). Assume that there exist $c \in \mathbb{R}_0^+$ such that $(\max_{i\in\{1,\ldots,n\}}\{n^{-1}R_i^{-6}\})_n \geq c$ infinitely often with non-zero probability. However, we have that for all fixed $k \in \mathbb{N}_0$, the sequence $(n^{-1}R_k^{-6})_n$ converges almost surely to 0 as $n \to \infty$. This entails that the (smallest) index $m(n)$ such that $n^{-1}R_{m(n)}^{-6} = \max_{i\in\{1,\ldots,n\}}\{n^{-1}R_i^{-6}\}$ satisfies $m(n) \to \infty$ as $n \to \infty$ with probability 1. Then, for $c \in \mathbb{R}_0^+$, and for any $N \in \mathbb{N}_0$ there exist with non-zero probability an $\underline{n} \geq N$ such that $\underline{n}^{-1}R_{\underline{n}}^{-6} \geq c$. Indeed, from previously stated arguments, there exist with non-zero probability a $\tilde{n}$ large enough allowing to get simultaneously

$$(i) \;\; c \leq \tilde{n}^{-1} R_{m(\tilde{n})}^{-6} \leq m(\tilde{n})^{-1} R_{m(\tilde{n})}^{-6}$$

and (ii) $\underline{n} := m(\tilde{n}) \geq N$. This is obviously a contradiction since the sequence $n^{-1} R_n^{-6}$ converges to 0 almost surely, as established earlier.

□

Proof of Theorem 4.7. Since the ARMA(l, q) process A_i, $i \in \mathbb{Z}$ is a purely non-deterministic zero-mean stable stationary process, the Wold Decomposition Theorem yields

$$A_i = \sum_{j=1}^{\infty} b_j \epsilon_{i-j} \tag{15}$$

with $\sum_{j=1}^{\infty} b_j^2 < \infty$ and $b_1 = 1$ - see for instance [7]. This directly entails that the A_i are identically distributed with finite variance. Now, we have by assumption on ϵ_1 that,

$$\mathrm{E}(R_i^{-6}) = \mathrm{E}(R_1^{-6}) = \mathrm{E}(e^{-6A_1}) = \mathrm{E}(e^{-6\sum_{j=1}^{\infty} b_j \epsilon_{1-j}}) = \prod_{j=1}^{\infty} \mathrm{E}(e^{-6 b_j \epsilon_1}) < \infty.$$

From now on, the rest of the proof follows exactly as in the proof of Theorem 4.6. □

The following technical result from [17] is required for the next proof.

Lemma A.1. *Let $T_n(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ a sequence of convex stochastic processes. Let $T(\tilde{\boldsymbol{\theta}})$ be a convex stochastic process such that the finite dimensional distributions of $T_n(\tilde{\boldsymbol{\theta}})$ converge to the finite dimensional distributions of $T(\tilde{\boldsymbol{\theta}})$ for all $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$. Let $\{\tilde{\boldsymbol{\theta}}_n, n \in \mathbb{N}\}$ the collection of random p-vectors such that for all $n \in \mathbb{N}_0$,*

$$\tilde{\boldsymbol{\theta}}_n = \inf_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} T_n(\tilde{\boldsymbol{\theta}}),$$

and $\tilde{\boldsymbol{\theta}}_{\lim}$ a random p-vector such that

$$\tilde{\boldsymbol{\theta}}_{\lim} = \inf_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} T(\tilde{\boldsymbol{\theta}}).$$

Then, we have that as $n \to \infty$, $\tilde{\boldsymbol{\theta}}_n \xrightarrow{\mathcal{L}} \tilde{\boldsymbol{\theta}}_{\lim}$.

Proof of Theorem 4.8. We consider without loss of generality and for the sake of simplicity that $\boldsymbol{\theta} = \mathbf{0}$ in this proof. It is well known that since $\boldsymbol{\theta} = \mathbf{0}$, $n^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(n)}$ minimizes (in $\tilde{\boldsymbol{\theta}}$) the following expression

$$\sum_{i=1}^{n} \|\mathbf{X}_i - n^{-1/2}\tilde{\boldsymbol{\theta}}\| - \|\mathbf{X}_i\| = \sum_{i=1}^{n} \|R_i \mathbf{V}^{1/2}\mathbf{U}_i - n^{-1/2}\tilde{\boldsymbol{\theta}}\| - \|R_i \mathbf{V}^{1/2}\mathbf{U}_i\|.$$

Using approximation (A3) in [17], we have that for all $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ and for all $0 < \delta < 1$,

$$
\begin{aligned}
&|\sum_{i=1}^{n} \|R_i \mathbf{V}^{1/2}\mathbf{U}_i - n^{-1/2}\tilde{\boldsymbol{\theta}}\| - \|R_i \mathbf{V}^{1/2}\mathbf{U}_i\| - n^{-1/2}\sum_{i=1}^{n} \frac{\mathbf{V}^{1/2}\mathbf{U}_i'}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|}\tilde{\boldsymbol{\theta}} \\
&-n^{-1}\sum_{i=1}^{n}(2R_i)^{-1}\|\mathbf{V}^{1/2}\mathbf{U}_i\|\tilde{\boldsymbol{\theta}}'\big(\mathbf{I}_p - \frac{\mathbf{V}^{1/2}\mathbf{U}_i\mathbf{U}_i'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\big)\tilde{\boldsymbol{\theta}}| \\
\leq \quad & C\, n^{-(2+\delta)/2}\sum_{i=1}^{n} R_i^{-(1+\delta)}\lambda_{\mathbf{V},1}^{1/2}\, \|\tilde{\boldsymbol{\theta}}\|^{2+\delta} \qquad (16)
\end{aligned}
$$

with $C \in \mathbb{R}_0^+$ some constant independent of $\tilde{\boldsymbol{\theta}}$ and $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and $\lambda_{\mathbf{V},1}$ the first eigenvalue of $\mathbf{V}$. Using the fact that $n^{-(2+\delta)/2}\sum_{i=1}^{n} R_i^{-(1+\delta)} = o_{\mathrm{P}}(1)$ for some $\delta \in (0,1)$, (16) yields, for all $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ as $n \to \infty$,

$$
\begin{aligned}
&\sum_{i=1}^{n} \|R_i \mathbf{V}^{1/2}\mathbf{U}_i - n^{-1/2}\tilde{\boldsymbol{\theta}}\| - \|R_i \mathbf{V}^{1/2}\mathbf{U}_i\| - n^{-1/2}\sum_{i=1}^{n} \frac{\mathbf{V}^{1/2}\mathbf{U}_i'}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|}\tilde{\boldsymbol{\theta}} \\
&-n^{-1}\sum_{i=1}^{n}(2R_i)^{-1}\tilde{\boldsymbol{\theta}}'\big(\mathbf{I}_p - \frac{\mathbf{V}^{1/2}\mathbf{U}_i\mathbf{U}_i'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\big)\tilde{\boldsymbol{\theta}} \\
= \quad & o_{\mathrm{P}}(1). \qquad (17)
\end{aligned}
$$

Assume now that $n^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(n)}$ is not $O_{\mathrm{P}}(1)$ and note that the collection

$$\{n^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(n)} \mid n = 1, \ldots, L\}$$

is bounded in probability for all fixed $L \in \mathbb{N}_0$. Then, there is ϵ in $(0,1)$ and a subsequence $\big(q(n)\big)^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(q(n))}$ such that for all $M \in \mathbb{R}_0^+$, there exist an $N \in \mathbb{N}$ such that

$$\mathrm{P}(\|\big(q(n)\big)^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(q(n))}\| \geq M) > \epsilon$$

for all $n \geq N$. Now, since $\big(q(n)\big)^{-1}\sum_{i=1}^{q(n)} R_i^{-1} = O_{\mathrm{P}}(1)$, we have by Prokhorov's Theorem that there exists a subsequence

$$\big(m(q(n))\big)^{-1}\sum_{i=1}^{m(q(n))}(2R_i)^{-1}\big(\mathbf{I}_p - \frac{\mathbf{V}^{1/2}\mathbf{U}_i\mathbf{U}_i'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\big)$$

converging in distribution to some random matrix $\mathbf{A}$. Obviously as $n \to \infty$ we also have that $\big(m(q(n))\big)^{-1/2}\sum_{i=1}^{m(q(n))}\frac{\mathbf{V}^{1/2}\mathbf{U}_i}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|} \xrightarrow{\mathcal{L}} \mathbf{Z}$, with $\mathbf{Z}$ multivariate Gaussian. Now, defining

$$T_{m(q(n))}(\tilde{\boldsymbol{\theta}}) = \sum_{i=1}^{m(q(n))} \|R_i \mathbf{V}^{1/2}\mathbf{U}_i - \big(m(q(n))\big)^{-1/2}\tilde{\boldsymbol{\theta}}\| - \|R_i\mathbf{V}^{1/2}\mathbf{U}_i\|$$

and

$$T(\tilde{\boldsymbol{\theta}}) = \mathbf{Z}'\tilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\theta}}'\mathbf{A}\,\tilde{\boldsymbol{\theta}},$$

the previous steps and (17) yields $T_{m(q(n))}(\tilde{\boldsymbol{\theta}}) \xrightarrow{\mathcal{L}} T(\tilde{\boldsymbol{\theta}})$ as $n \to \infty$. Now, using Lemma A.1 with $T_{m(q(n))}(\tilde{\boldsymbol{\theta}})$ and $T(\tilde{\boldsymbol{\theta}})$ we get that the sequence

$$\big(m(q(n))\big)^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(m(q(n)))}$$

is asymptotically distributed as the minimizer of $T(\tilde{\boldsymbol{\theta}})$ and then satisfies as $n \to \infty$,

$$\mathbf{Z} + \mathbf{A}\,\big(m(q(n))\big)^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(m(q(n)))} \quad = \quad o_{\mathrm{P}}(1). \tag{18}$$

It is easy to check that

$$\mathbf{A} \overset{\mathcal{L}}{=} \lim_{n\to\infty} \big(m(q(n))\big)^{-1} \sum_{i=1}^{m(q(n))} (2R_i)^{-1}\big(\mathbf{I}_p - \frac{\mathbf{V}^{1/2}\mathbf{U}_i\mathbf{U}_i'\mathbf{V}^{1/2}}{\|\mathbf{V}^{1/2}\mathbf{U}_i\|^2}\big)$$

is positive definite with probability 1 because by assumption there is no subsequence such that $s(n)^{-1}\sum_{i=1}^{s(n)} R_i^{-1}$ is $o_{\mathrm{P}}(1)$. Then, (18) yields, as $n \to \infty$,

$$\big(m(q(n))\big)^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(m(q(n)))} \quad = \quad O_{\mathrm{P}}(1). \tag{19}$$

There is a contradiction between (19) and the fact that for all $M \in \mathbb{R}_0$, there exist by assumption an $N \in \mathbb{N}$ such that $\mathrm{P}(\|\big(q(n)\big)^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(q(n))}\| \geq M) \geq \epsilon$ for every $n \geq N$ and for some $\epsilon > 0$. Then, we must have $n^{1/2}\hat{\boldsymbol{\theta}}_{\mathrm{sign}}^{(n)} = O_{\mathrm{P}}(1)$ as $n \to \infty$. This concludes the proof.

□

Proof of Theorem 5.1. It is easy to check that the fact that $\mathbf{V}^{(n)}$ depends on n does not change anything in the proofs of Theorems 4.1 and 4.3. Then, by Theorems 4.1 and 4.3, we have that $S^{(n)}(\boldsymbol{\theta})$ and $S^{(n)}(\hat{\boldsymbol{\theta}}^{(n)})$ have the same asymptotic distribution under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V}^{(n)},R_1,\ldots,R_n}$ as $n \to \infty$.

Now, the asymptotic distribution of $S^{(n)}(\boldsymbol{\theta})$ has been derived in scenario I under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\mathbf{V}^{(n)},R_1,\ldots,R_n}$ by [6]. This yields the result in scenarios II and III since $S^{(n)}(\boldsymbol{\theta})$ is measurable with respect to $\mathbf{U}_1,\ldots,\mathbf{U}_n$.

□

Proof of Theorem 5.3. Recall here that in this proof, we have that $\mathbf{V} = \mathbf{I}_p$. We keep the notation $\mathbf{V}$ to make explicit when the shape parameter appears in the various expressions. We also write $\mathbf{V}^{(n)} = \mathbf{V} + n^{-1/2}\boldsymbol{\tau}^{(n)}$. We will prove that for any $\epsilon \in \mathbb{R}_0^+$,

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}\Bigg[|\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) \\ -\big(\mathrm{v\overset{\circ}{e}ch}\,'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V}) \; - \frac{1}{2}\mathrm{v\overset{\circ}{e}ch}\,'(\boldsymbol{\tau})\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{V})\mathrm{v\overset{\circ}{e}ch}\,(\boldsymbol{\tau})\big)| > \epsilon\Bigg] = 0. \tag{20}$$

and that

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}\big[|\mathrm{v\overset{\circ}{e}ch}\,'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) - \mathrm{v\overset{\circ}{e}ch}\,'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V})| > \epsilon\big] \;=\; 0. \tag{21}$$

Combining (20) and (21) and using the continuous mapping theorem will yield the result. We first prove (20). We note first that by definition of f_n,

$$\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) \;=\; \frac{(p+\nu_n)}{2}\sum_{i=1}^{n}\Bigg(\log(1+\frac{(d_{ni}(\boldsymbol{\theta},\mathbf{V})/\sigma)^2}{\nu_n}) \\ -\log(1+\frac{(d_{ni}(\boldsymbol{\theta},\mathbf{V}^{(n)})/\sigma)^2}{\nu_n})\Bigg). \tag{22}$$

Consider the representation

$$\mathbf{X}_{n1} = R_{n1}\sigma\mathbf{V}^{1/2}\mathbf{U}_{n1} + \boldsymbol{\theta}, \ldots, \mathbf{X}_{nn} = R_{nn}\sigma\mathbf{V}^{1/2}\mathbf{U}_{nn} + \boldsymbol{\theta}$$

as in (1). We can rewrite (22) as

$$\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) \;=\; \frac{(p+\nu_n)}{2}\sum_{i=1}^{n}\big(\log(R_{ni}^{-2}\nu_n + \mathbf{U}'_{ni}\mathbf{V}^{-1}\mathbf{U}_{ni}) \\ -\log(R_{ni}^{-2}\nu_n + \mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni})\big). \tag{23}$$

Recalling that $\mathbf{V} = \mathbf{I}_p$, simple Taylor expansion of the logarithm function around $(R_{ni}^{-2}\nu_n + 1)$ yields

$$\begin{aligned}\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) &= -\frac{(p+\nu_n)}{2}\sum_{i=1}^{n}\Big(\log(R_{ni}^{-2}\nu_n+1)\\ &\quad -\log(R_{ni}^{-2}\nu_n+\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni})\Big)\\ &= \frac{(p+\nu_n)}{2}\sum_{i=1}^{n}\Big((R_{ni}^{-2}\nu_n+1)^{-1}(\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}-1)\\ &\quad -\frac{1}{2}(R_{ni}^{-2}\nu_n+1)^{-2}(\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}-1)^2\\ &\quad +\frac{1}{3}(H_{ni})^{-3}(\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}-1)^3\Big)\end{aligned} \tag{24}$$

for some

$$H_{ni}\in[R_{ni}^{-2}\nu_n+\min(1,\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}),R_{ni}^{-2}\nu_n+\max(1,\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni})].$$

It is not hard to see that since $R_{ni}^{-2}\nu_n$ is positive and $\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}$ is almost surely positive and bounded away from 0 (this is easy to deduce from the structure of $\mathbf{V}^{(n)}$), we have that $(H_{ni})^{-3}=O_{\mathrm{P}}(1)$. Now, write the spectral decomposition

$$(\mathbf{V}^{(n)})^{-1}=(\boldsymbol{\beta}^{(n)})'(\boldsymbol{\Lambda}^{(n)})^{-1}\boldsymbol{\beta}^{(n)}=:(\boldsymbol{\beta}^{(n)})'(\mathbf{I}_p+n^{-1/2}\boldsymbol{l}^{(n)})^{-1}\boldsymbol{\beta}^{(n)},$$

with $\boldsymbol{l}^{(n)}$ a diagonal matrix and $\boldsymbol{\beta}^{(n)}\in\mathcal{SO}_p$ such that $(\boldsymbol{\beta}^{(n)})'\boldsymbol{l}^{(n)}\boldsymbol{\beta}^{(n)}=\boldsymbol{\tau}^{(n)}$. Using a Taylor expansion we get that

$$\begin{aligned}\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}-1 &= \mathbf{U}'_{ni}\boldsymbol{\beta}^{(n)}(\boldsymbol{\Lambda}^{(n)})^{-1}(\boldsymbol{\beta}^{(n)})'\mathbf{U}_{ni}-1\\ &= \mathbf{U}'_{ni}\boldsymbol{\beta}^{(n)}(\mathbf{I}_p-n^{-1/2}\boldsymbol{l}^{(n)}+n^{-1}(\boldsymbol{l}^{(n)})^2)(\boldsymbol{\beta}^{(n)})'\mathbf{U}_{ni}-1\\ &\quad +O_{\mathrm{P}}(n^{-3/2}).\\ &= \mathbf{U}'_{ni}(-n^{-1/2}\boldsymbol{\tau}^{(n)}+n^{-1}(\boldsymbol{\tau}^{(n)})^2)\mathbf{U}_{ni}+O_{\mathrm{P}}(n^{-3/2}).\end{aligned} \tag{25}$$

This immediately yields $(\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}-1)^3=O_{\mathrm{P}}(n^{-3/2})$. Combining this last fact and $(H_{ni})^{-3}=O_{\mathrm{P}}(1)$ yields, as $n\to\infty$,

$$\begin{aligned}\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) &= -\frac{(p+\nu_n)}{2}\sum_{i=1}^{n}\big((R_{ni}^{-2}\nu_n+1)^{-1}(\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}-1)\\ &\quad -\frac{1}{2}(R_{ni}^{-2}\nu_n+1)^{-2}(\mathbf{U}'_{ni}(\mathbf{V}^{(n)})^{-1}\mathbf{U}_{ni}-1)^2\big)+o_{\mathrm{P}}(1).\end{aligned} \tag{26}$$

Combining (25) with (26) and using that $(R_{ni}^{-2}\nu_n+1)^{-1}$ is almost surely bounded, we get

$$\begin{aligned}\log\left(\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) &= \frac{(p+\nu_n)}{2}\Bigg(n^{-1/2}\sum_{i=1}^{n}(R_{ni}^{-2}\nu_n+1)^{-1}\mathbf{U}'_{ni}\boldsymbol{\tau}^{(n)}\mathbf{U}_{ni}\\ &\quad +n^{-1}\sum_{i=1}^{n}\frac{1}{2}(R_{ni}^{-2}\nu_n+1)^{-2}(\mathbf{U}'_{ni}\boldsymbol{\tau}^{(n)}\mathbf{U}_{ni})^2\\ &\quad -(R_{ni}^{-2}\nu_n+1)^{-1}(\mathbf{U}'_{ni}(\boldsymbol{\tau}^{(n)})^2\mathbf{U}_{ni})\Bigg)+o_{\mathrm{P}}(1).\\ &= \frac{(p+\nu_n)}{2}\Bigg(n^{-1/2}\sum_{i=1}^{n}(R_{ni}^{-2}\nu_n+1)^{-1}\mathrm{tr}(\boldsymbol{\tau}^{(n)}\mathbf{U}_{ni}\mathbf{U}'_{ni})\\ &\quad +n^{-1}\sum_{i=1}^{n}[\frac{1}{2}(R_{ni}^{-2}\nu_n+1)^{-2}\mathrm{tr}^2(\boldsymbol{\tau}^{(n)}\mathbf{U}_{ni}\mathbf{U}'_{ni})\\ &\quad -(R_{ni}^{-2}\nu_n+1)^{-1}\mathrm{tr}(\boldsymbol{\tau}^{(n)}\mathbf{U}_{ni}\mathbf{U}'_{ni}\boldsymbol{\tau}^{(n)})]\Bigg)+o_{\mathrm{P}}(1).\\ &= \frac{(p+\nu_n)}{2}\Bigg(n^{-1/2}\sum_{i=1}^{n}\big[(R_{ni}^{-2}\nu_n+1)^{-1}\\ &\quad \mathrm{vec}(\boldsymbol{\tau}^{(n)})'\mathrm{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni})\big]\\ &\quad +n^{-1}\sum_{i=1}^{n}\big[\frac{1}{2}(R_{ni}^{-2}\nu_n+1)^{-2}\\ &\quad \mathrm{vec}(\boldsymbol{\tau}^{(n)})'\mathrm{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni})\mathrm{vec}'(\mathbf{U}_{ni}\mathbf{U}'_{ni})\mathrm{vec}(\boldsymbol{\tau}^{(n)})\\ &\quad -(R_{ni}^{-2}\nu_n+1)^{-1}\mathrm{vec}(\boldsymbol{\tau}^{(n)}\mathbf{U}_{ni}\mathbf{U}'_{ni})'\mathrm{vec}(\boldsymbol{\tau}^{(n)})\big]\Bigg)\\ &\quad +o_{\mathrm{P}}(1).\end{aligned} \tag{27}$$

Recall now that the definition of f_n implies that under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$, $(\nu_n)^{-1/2}\mathbf{X}_{1n}$ admits density

$$f_{(\nu_n)^{-1/2}\mathbf{X}_{1n}}(\mathbf{x})=\frac{\Gamma((p+\nu_n)/2)}{\pi^{p/2}\sigma^p\,\Gamma(\nu_n/2)}(1+\frac{(\mathbf{x}-\boldsymbol{\theta})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\theta})}{\sigma^2})^{-\frac{p+\nu_n}{2}},$$

which implies that $(\nu_n)^{-1/2}R_{n1}$ admits density

$$f_{(\nu_n)^{-1/2}R_{1n}}(r)=\frac{\Gamma((p+\nu_n)/2)}{\pi^{p/2}\sigma^p\,\Gamma(\nu_n/2)}r^{p-1}(1+r^2/\sigma^2)^{-\frac{p+\nu_n}{2}}.$$

It is easy to see from this last fact that for all $C\in\mathbb{R}^+$ and as $n\to\infty$, we have that $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}\big((\nu_n)^{-1/2}R_{1n})\leq C\big)\to 0$. This obviously directly entails that for all $i \in \{1,\ldots,n\}$ we have $(R_{ni}^{-2}\nu_n+1)^{-1} = 1+o_{\mathrm{P}}(1)$ as $n\to\infty$. Combining this fact with (27) and using the Law of Large Numbers yields, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$ as $n\to\infty$,

$$\begin{aligned}
\log\left(\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) &= \frac{(p+\nu_n)}{2}\Bigg(n^{-1/2}\sum_{i=1}^{n}(R_{ni}^{-2}\nu_n+1)^{-1} \\
&\quad \mathrm{vec}'(\boldsymbol{\tau}^{(n)})\mathrm{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni}) \\
&\quad +\frac{1}{2}\mathrm{vec}(\boldsymbol{\tau}^{(n)})'\mathrm{E}(\mathrm{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni})\mathrm{vec}'(\mathbf{U}_{ni}\mathbf{U}'_{ni}))\mathrm{vec}(\boldsymbol{\tau}^{(n)}) \\
&\quad -\mathrm{vec}(\boldsymbol{\tau}^{(n)}\mathrm{E}(\mathbf{U}_{ni}\mathbf{U}'_{ni}))'\mathrm{vec}(\boldsymbol{\tau}^{(n)})\Bigg)+o_{\mathrm{P}}(1) \\
&= \frac{(p+\nu_n)}{2}\Bigg(\frac{n^{-1/2}}{p+\nu_n}\sum_{i=1}^{n}\big(-\frac{\dot{f}_n}{f_n}(R_{ni})R_{ni} \\
&\quad \mathrm{vec}(\boldsymbol{\tau}^{(n)})\mathrm{vec}(\mathbf{U}_{ni}\mathbf{U}'_{ni})\big) \\
&\quad +\frac{1}{2p(p+2)}\mathrm{vec}(\boldsymbol{\tau}^{(n)})'\big(\mathbf{I}_{p^2}+\mathbf{K}_p+\mathrm{vec}(\mathbf{I}_p) \\
&\quad \mathrm{vec}'(\mathbf{I}_p)\big)\mathrm{vec}(\boldsymbol{\tau}^{(n)}) \\
&\quad -\frac{1}{2p}\mathrm{vec}(\boldsymbol{\tau}^{(n)})'(\mathbf{I}_{p^2}+\mathbf{K}_p)\mathrm{vec}(\boldsymbol{\tau}^{(n)})\Bigg)+o_{\mathrm{P}}(1).
\end{aligned}$$

The last line is a well known fact that can be found in [6]. Now, using that

$$\begin{aligned}
0 &= -\frac{n(p+\nu_n)}{2p}\big(\log(\det(\mathbf{V}+n^{-1/2}\boldsymbol{\tau}^{(n)}))-\log(\det(\mathbf{V}))\big) \\
&= -\frac{n^{1/2}(p+\nu_n)}{2p}\mathrm{tr}(\boldsymbol{\tau}^{(n)})+\frac{p+\nu_n}{4p}\mathrm{tr}((\boldsymbol{\tau}^{(n)})^2)+o(1),
\end{aligned}$$

we get that as $n\to\infty$ and under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$,

$$\begin{aligned}
\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) &= \frac{(p+\nu_n)}{2}\Bigg(\mathrm{vec}(\boldsymbol{\tau}^{(n)})'\big(\frac{n^{-1/2}}{p+\nu_n}\sum_{i=1}^n -\frac{\dot{f}_n}{f_n}(R_{ni})R_{ni} \\
&\quad \mathrm{vec}(\mathbf{U}_{ni}\mathbf{U}_{ni}') - \frac{n^{1/2}}{p}\mathrm{vec}(\mathbf{I}_p)\big) \\
&\quad +\frac{1}{2p(p+2)}\mathrm{vec}(\boldsymbol{\tau}^{(n)})'\big(\mathbf{I}_{p^2}+\mathbf{K}_p+\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}'(\mathbf{I}_p)\big) \\
&\quad \mathrm{vec}(\boldsymbol{\tau}^{(n)}) \\
&\quad -\frac{1}{2p}\mathrm{vec}(\boldsymbol{\tau}^{(n)})'(\mathbf{I}_{p^2}+\mathbf{K}_p)\mathrm{vec}(\boldsymbol{\tau}^{(n)}) \\
&\quad +\frac{1}{4p}\mathrm{vec}(\boldsymbol{\tau}^{(n)})'(\mathbf{I}_{p^2}+\mathbf{K}_p)\mathrm{vec}(\boldsymbol{\tau}^{(n)})\Bigg) + o_{\mathrm{P}}(1). \\
&= \mathrm{vec}'(\boldsymbol{\tau}^{(n)})\frac{n^{1/2}}{2}\mathrm{vec}(\mathbf{S}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V})-\mathbf{I}_p) \\
&\quad -\frac{1}{2}\times\frac{p}{4(p+2)} \\
&\quad \mathrm{vec}(\boldsymbol{\tau}^{(n)})'\big(\mathbf{I}_{p^2}+\mathbf{K}_p-\frac{2}{p}\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}'(\mathbf{I}_p)\big)\mathrm{vec}(\boldsymbol{\tau}^{(n)}) \\
&\quad +o_{\mathrm{P}}(1).
\end{aligned} \tag{28}$$

Now, it is easy to show that as $n\to\infty$,

$$\mathbf{M}_p'(\mathrm{v\overset{\circ}{e}ch}\,(\boldsymbol{\tau}^{(n)})) = \mathrm{vec}(\boldsymbol{\tau}^{(n)}) + o(1),$$

which, combined with (28) and basic properties of $\mathbf{M}_p'$, immediately yields (20) (recall here once again that $\mathbf{V}=\mathbf{I}_p$). We now prove (21). We will show that as $n\to\infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$,

$$\boldsymbol{\Delta}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V}) = \boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) + o_{\mathrm{P}}(1) \tag{29}$$

It is obviously sufficient to show that as $n\to\infty$ and under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$,

$$n^{1/2}\mathbf{S}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V}) - n^{1/2}\mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) = o_{\mathrm{P}}(1). \tag{30}$$

First, we have that under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$,

$$\begin{aligned}
n\,\mathrm{E}(\|\mathbf{S}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V})-\mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V})\|^2_{\mathrm{F}}) &= n^{-1}\mathrm{E}(\|\sum_{i=1}^n\big((p+\nu_n)\frac{1}{1+\frac{d_{ni}(\boldsymbol{\theta},\mathbf{V})^2}{\sigma^2\nu_n}} \\
&\quad \frac{d_{ni}(\boldsymbol{\theta},\mathbf{V})^2}{\sigma^2\nu_n}\mathbf{U}_{ni}(\boldsymbol{\theta},\mathbf{V})\mathbf{U}_{ni}(\boldsymbol{\theta},\mathbf{V})' \\
&\quad -p\,\mathbf{U}_{ni}(\boldsymbol{\theta},\mathbf{V})\mathbf{U}_{ni}(\boldsymbol{\theta},\mathbf{V})'\big)\|^2_{\mathrm{F}}).
\end{aligned}$$

We write

$$\begin{aligned} c_{ni} &:= (p+\nu_n)(1+\frac{d_{ni}(\boldsymbol{\theta},\mathbf{V})^2}{\sigma^2\nu_n})^{-1}\frac{d_{ni}(\boldsymbol{\theta},\mathbf{V})^2}{\sigma^2\nu_n} - p \\ &= -\frac{\dot{f}_n}{f_n}(d_{ni}(\boldsymbol{\theta},\mathbf{V})/\sigma)d_{ni}(\boldsymbol{\theta},\mathbf{V})/\sigma - p. \end{aligned}$$

Letting $k_p^{(n)}$ be a normalization constant, we have that under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$ the density of the $d_{ni}(\boldsymbol{\theta},\mathbf{V})/\sigma$'s is $k_p^{(n)} r^{p-1} f_n(r)\mathbb{I}_{\{r>0\}}$. This yields

$$\mathrm{E}(c_{ni}) = -k_p^{(n)}\int_0^\infty \dot{f}_n(r) r^p \mathrm{d}r - p = 0$$

by integration by part. Similarly,

$$J_p(f_n) = k_p^{(n)}\int_0^\infty r^{p+1}(\frac{\dot{f}_n}{f_n})^2(r) f_n(r)\mathrm{d}r = \mathrm{E}((c_{ni}+p)^2) = \mathrm{E}(c_{ni}^2) + p^2.$$

With these notations, we write

$$\begin{aligned} & n\,\mathrm{E}(\|\mathbf{S}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V}) - \mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V})\|^2_{\mathrm{F}}) \\ = \; & n^{-1}\mathrm{E}(\mathrm{tr}(\sum_{i=1}^n c_{ni}\mathbf{U}_{ni}(\boldsymbol{\theta},\mathbf{V})\mathbf{U}'_{ni}(\boldsymbol{\theta},\mathbf{V})\sum_{j=1}^n c_{nj}\mathbf{U}_{nj}(\boldsymbol{\theta},\mathbf{V})\mathbf{U}'_{nj}(\boldsymbol{\theta},\mathbf{V}))). \end{aligned}$$

Using $\mathrm{E}(c_{ni}) = 0$ and independence, we get that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$,

$$\begin{aligned} & n\,\mathrm{E}(\|\mathbf{S}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V}) - \mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V})\|^2_{\mathrm{F}}) \\ = \; & n^{-1}\mathrm{E}(\sum_{i=1}^n c_{ni}^2\mathrm{tr}(\mathbf{U}_{ni}(\boldsymbol{\theta},\mathbf{V})\mathbf{U}'_{ni}(\boldsymbol{\theta},\mathbf{V})\mathbf{U}_{ni}(\boldsymbol{\theta},\mathbf{V})\mathbf{U}'_{ni}(\boldsymbol{\theta},\mathbf{V}))) \\ = \; & n^{-1}\sum_{i=1}^n \mathrm{E}(c_{ni}^2) \end{aligned}$$

Now, since under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}$,

$$\begin{aligned} & \lim_{n\to\infty} n\,\mathrm{E}(\|\mathbf{S}^{(n)}_{f_n}(\boldsymbol{\theta},\sigma,\mathbf{V}) - \mathbf{S}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V})\|^2_{\mathrm{F}}) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathrm{E}(c_{ni}^2) \\ & = \lim_{n\to\infty}\mathrm{E}(c_{n1}^2) = \lim_{n\to\infty} J_p(f_n) - p^2 = \lim_{n\to\infty}\frac{p(p+2)(p+\nu_n)}{(p+2+\nu_n)} - p^2 = 0, \end{aligned}$$

we have that (29) holds (a closed form for $J_p(f_n)$ can be found for instance in [6]), which means that (21) holds. Combining (20) and (21) and using the continuous mapping theorem concludes the proof.

As a closing word, we note that the fact that $\mathbf{V} = \mathbf{I}_p$ has been used solely for the sake of simplifying the various expressions and computation but does not play a fundamental role. The proof would in fact hold for any shape matrix modulo some small adjustments.

□

Proof of Theorem 5.5. It is sufficient to prove the following LAN property,

$$\lim_{n\to\infty} \tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}\Bigg[\Big|\log\left(\frac{\mathrm{d}\tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},\tilde{f}_n}}{\mathrm{d}\tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}}\right) - \Big(\mathrm{ve\overset{\circ}{c}h}'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) - \frac{1}{2}\mathrm{ve\overset{\circ}{c}h}'(\boldsymbol{\tau})\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{V})\mathrm{ve\overset{\circ}{c}h}\,(\boldsymbol{\tau})\Big)\Big| > \epsilon\Bigg] = 0. \quad (31)$$

Now, writing

$$T_n(\mathbf{X}_1,\ldots,\mathbf{X}_n) = \Big|\log\left(\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}}\right) - \Big(\mathrm{ve\overset{\circ}{c}h}'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) - \frac{1}{2}\mathrm{ve\overset{\circ}{c}h}'(\boldsymbol{\tau})\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{V})\mathrm{ve\overset{\circ}{c}h}\,(\boldsymbol{\tau})\Big)\Big|,$$

writing $g_n(r_1,\ldots,r_n) := \prod_{i=1}^n f_n(r_i)$ and using $\left\|\frac{\tilde{f}_n}{g_n}(r_1,\ldots,r_n)\right\|_\infty = 1 + o(1)$, we have as $n \to \infty$ that

$$
\begin{aligned}
& \tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}\Bigg[\Big| \log\left(\frac{\mathrm{d}\tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},\tilde{f}_n}}{\mathrm{d}\tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}} \right) \\
& -\big(\overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) - \frac{1}{2}\overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{V})\overset{\circ}{\mathrm{vech}}(\boldsymbol{\tau})\big)\Big| > \epsilon \Bigg] \\
= \; & \tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}\Bigg[\Big| \log\left(\frac{\mathrm{d}\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{d}\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}} \right) + \log(1+o_{\mathrm{P}}(1)) \\
& -\big(\overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) - \frac{1}{2}\overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{V})\overset{\circ}{\mathrm{vech}}(\boldsymbol{\tau})\big)\Big| > \epsilon \Bigg] \\
= \; & \tilde{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},\tilde{f}_n}\Bigg[\Big| \log\left(\frac{\mathrm{d}\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V}^{(n)},f_n}}{\mathrm{d}\mathrm{P}^{(n)}_{\boldsymbol{\theta},\sigma,\mathbf{V},f_n}} \right) \\
& -\big(\overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\boldsymbol{\Delta}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta},\mathbf{V}) - \frac{1}{2}\overset{\circ}{\mathrm{vech}}{}'(\boldsymbol{\tau})\boldsymbol{\Gamma}_{\mathrm{sign}}(\mathbf{V})\overset{\circ}{\mathrm{vech}}(\boldsymbol{\tau})\big)\Big| > \epsilon \Bigg] + o(1) \\
= \; & \int_{\mathbb{R}^{n\times p}} \mathbb{I}_{\{T_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)>\epsilon\}}(\mathbf{x}_1,\ldots,\mathbf{x}_n)\tilde{f}_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)\mathrm{d}\mathbf{x}_1\ldots\mathrm{d}\mathbf{x}_n + o(1) \\
= \; & \int_{\mathbb{R}^{n\times p}} \mathbb{I}_{\{T_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)>\epsilon\}}(\mathbf{x}_1,\ldots,\mathbf{x}_n)g_n(\mathbf{x}_1,\ldots,\mathbf{x}_n)(1+o(1))\mathrm{d}\mathbf{x}_1\ldots\mathrm{d}\mathbf{x}_n \\
& +o(1) \\
= \; & o(1).
\end{aligned}
$$

We used Theorem 5.3 in the last line. It follows that the LAN property (31) holds and the result is a consequence of classical Le Cam asymptotic theory of experiments.

□

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 811017.

References

[1] Croux, C., Ollila, E. and Oja, H. (2002). Sign and rank covariance matrices: statistical properties and application to principal components analysis. In *Statistical data analysis based on the L1-norm and related methods* 257–269. Springer.

[2] Cuesta-Albertos, J. A., Cuevas, A. and Fraiman, R. (2009). On projection-based tests for directional and compositional data. *Statist. Comput.* **19** 367–380.

[3] Dürre, A., Tyler, D. E. and Vogel, D. (2014). The spatial sign covariance matrix with unknown location. *J. Multivariate Anal.* **130** 107–117.
[4] Francq, C., Jiménez-Gamero, M. and Meintanis, S. (2017). Tests for conditional ellipticity in multivariate GARCH models. *J. Econometrics* **196** 305–319.
[5] Groeneveld, R. A. (1991). An influence function approach to describing the skewness of a distribution. *Amer. Statist.* **45** 97–102.
[6] 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.
[7] Hamilton, J. D. (1994). *Time Series Analysis*, 1st edition ed. Princeton University Press, Princeton.
[8] Henze, N., Hlávka, Z. and Meintanis, S. G. (2014). Testing for spherical symmetry via the empirical characteristic function. *Statistics* **48** 1282–1296.
[9] John, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127.
[10] John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173.
[11] Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*, 1st ed. Springer, New York.
[12] Le Cam, L. and Yang, G. L. (2000). *Asymptotics in statistics*, 2nd ed. Springer, New York.
[13] Li, Z. and Yao, J. (2016). Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size. *Electron. J. Statist.* **10** 2973–3010.
[14] Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Statist.* **11** 204–209.
[15] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Statist.* **5** 201–213.
[16] Muirhead, R. J. and Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika* **67** 31–43.
[17] Möttönen, J., Nordhausen, K. and Oja, H. (2010). Asymptotic theory of the spatial median. *Nonparametrics and robustness in modern statistical inference and time series analysis: A Festchrift in honor of Professor Jana Jurečková* **7** 182–193.
[18] Onatski, A., Moreira, M. J. and Hallin, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231.
[19] Onatski, A., Moreira, M. J. and Hallin, M. (2014). Signal detection in high dimension: the multispiked case. *Ann. Statist.* **42** 225–254.
[20] Paindaveine, D. (2008). A canonical definition of shape. *Statist. Probab. Lett.* **78** 2240–2247.
[21] Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli* **22** 1745–1769.
[22] Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045–1050.

[23] Sen, B. and Huang, Z. (2023). Multivariate symmetry: distribution-free testing via optimal transport. *arXiv preprint arXiv:2305.01839*.

[24] Sirkiä, S., Taskinen, S., Oja, H. and Tyler, D. E. (2009). Tests and estimates of shape based on spatial signs and ranks. *J. Nonparametr. Statist.* **21** 155–176.

[25] Tian, X., Lu, Y. and Li, W. (2015). A robust test for sphericity of high-dimensional covariance matrices. *J. Multivariate Anal.* **141** 217–227.

[26] Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika* **69** 429–436.

[27] Tyler, D. E. (1987). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika* **74** 579–589.