

TESTING FOR PRINCIPAL COMPONENT DIRECTIONS UNDER WEAK IDENTIFIABILITY

BY DAVY PAINDAVEINE^{*}, JULIEN REMY^{**} AND THOMAS VERDEBOUT[†]

ECARES and Département de Mathématique, Université libre de Bruxelles, ^{*}dpaindav@ulb.ac.be; ^{**}Julien.Remy@vub.ac.be; [†]tverdebo@ulb.ac.be

We consider the problem of testing, on the basis of a p -variate Gaussian random sample, the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ against the alternative $\mathcal{H}_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0$, where $\boldsymbol{\theta}_1$ is the “first” eigenvector of the underlying covariance matrix and $\boldsymbol{\theta}_1^0$ is a fixed unit p -vector. In the classical setup where eigenvalues $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_p$ are fixed, the Anderson (*Ann. Math. Stat.* **34** (1963) 122–148) likelihood ratio test (LRT) and the Hallin, Paindaveine and Verdebout (*Ann. Statist.* **38** (2010) 3245–3299) Le Cam optimal test for this problem are asymptotically equivalent under the null hypothesis, hence also under sequences of contiguous alternatives. We show that this equivalence does not survive asymptotic scenarios where $\lambda_{n1}/\lambda_{n2} = 1 + O(r_n)$ with $r_n = O(1/\sqrt{n})$. For such scenarios, the Le Cam optimal test still asymptotically meets the nominal level constraint, whereas the LRT severely over-rejects the null hypothesis. Consequently, the former test should be favored over the latter one whenever the two largest sample eigenvalues are close to each other. By relying on the Le Cam’s asymptotic theory of statistical experiments, we study the non-null and optimality properties of the Le Cam optimal test in the aforementioned asymptotic scenarios and show that the null robustness of this test is not obtained at the expense of power. Our asymptotic investigation is extensive in the sense that it allows r_n to converge to zero at an arbitrary rate. While we restrict to single-spiked spectra of the form $\lambda_{n1} > \lambda_{n2} = \dots = \lambda_{np}$ to make our results as striking as possible, we extend our results to the more general elliptical case. Finally, we present an illustrative real data example.

1. Introduction. Principal Component Analysis (PCA) is one of the most classical tools in multivariate statistics. For a random p -vector \mathbf{X} with mean zero and a covariance matrix $\boldsymbol{\Sigma}$ admitting the spectral decomposition $\boldsymbol{\Sigma} = \sum_{j=1}^p \lambda_j \boldsymbol{\theta}_j \boldsymbol{\theta}_j'$ ($\lambda_1 \geq \dots \geq \lambda_p$), the j th principal component is $\boldsymbol{\theta}_j' \mathbf{X}$, that is, the projection of \mathbf{X} onto the j th unit eigenvector $\boldsymbol{\theta}_j$ of $\boldsymbol{\Sigma}$. In practice, $\boldsymbol{\Sigma}$ is usually unknown, so that one of the key issues in PCA is to perform inference on eigenvectors. The seminal paper Anderson (1963) focused on the multinormal case and derived asymptotic results for the maximum likelihood estimators of the $\boldsymbol{\theta}_j$ ’s and λ_j ’s. Later, Tyler (1981, 1983) extended those results to the elliptical case, where, to avoid moment assumptions, $\boldsymbol{\Sigma}$ is then the corresponding “scatter” matrix rather than the covariance matrix. Still under ellipticity assumptions, Hallin, Paindaveine and Verdebout (2010) obtained Le Cam optimal tests on eigenvectors and eigenvalues, whereas Hallin, Paindaveine and Verdebout (2014) developed efficient R-estimators for eigenvectors. Croux and Haesbroeck (2000), Hubert, Rousseeuw and Vanden Branden (2005) and He et al. (2011) proposed various robust methods for PCA. Recently, Johnstone and Lu (2009), Berthet and Rigollet (2013) and Han and Liu (2014) considered inference on eigenvectors of $\boldsymbol{\Sigma}$ in sparse high-dimensional

Received October 2017; revised July 2018.

MSC2010 subject classifications. Primary 62F05, 62H25; secondary 62E20.

Key words and phrases. Le Cam’s asymptotic theory of statistical experiments, local asymptotic normality, principal component analysis, spiked covariance matrices, weak identifiability.

situations. PCA has also been extensively considered in the functional case; see, for example, Boente and Fraiman (2000), Bali et al. (2011) or the review paper Cuevas (2014).

In this work, we focus on the problem of testing the null hypothesis $\mathcal{H}_0 : \theta_1 = \theta_1^0$ against the alternative $\mathcal{H}_1 : \theta_1 \neq \theta_1^0$, where θ_1^0 is a given unit vector of \mathbb{R}^p . While, strictly speaking, the fact that θ_1 will below be defined up to a sign only should lead us to formulate the null hypothesis as $\mathcal{H}_0 : \theta_1 \in \{\theta_1^0, -\theta_1^0\}$, we will stick to the formulation above, which is the traditional one in the literature; we refer to the many references provided below. We restrict to θ_1 for the sake of simplicity only; our results could indeed be extended to null hypotheses of the form $\mathcal{H}_0 : \theta_j = \theta_j^0$ for any other j . While the emphasis in PCA is usually more on point estimation, the testing problems above are also of high practical relevance. For instance, they are of paramount importance in *confirmatory PCA*, that is, when it comes to testing that θ_1 (or any other θ_j) coincides with an eigenvector obtained from an earlier real data analysis (“historical data”) or with an eigenvector resulting from a theory or model. In line with this, tests for the null hypothesis $\mathcal{H}_0 : \theta_1 = \theta_1^0$ have been used in, among others, Jackson (2005) to analyze the concentration of a chemical component in a solution and in Sylvester, Kramer and Jungers (2008) for the study of the geometric similarity in modern humans.

More specifically, we want to test $\mathcal{H}_0 : \theta_1 = \theta_1^0$ against $\mathcal{H}_1 : \theta_1 \neq \theta_1^0$ on the basis of a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from the p -variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (the extension to elliptical distributions will also be considered). Denoting as $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$ the eigenvalues of the sample covariance matrix $\mathbf{S} := \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$ (as usual, $\bar{\mathbf{X}} := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$ here), the classical test for this problem is the Anderson (1963) likelihood ratio test, ϕ_A say, rejecting the null hypothesis at asymptotic level α when

$$Q_A := n(\hat{\lambda}_1 \theta_1^{0'} \mathbf{S}^{-1} \theta_1^0 + \hat{\lambda}_1^{-1} \theta_1^{0'} \mathbf{S} \theta_1^0 - 2) > \chi_{p-1, 1-\alpha}^2,$$

where $\chi_{\ell, 1-\alpha}^2$ stands for the α -upper quantile of the chi-square distribution with ℓ degrees of freedom. Various extensions of this test have been proposed in the literature: to mention only a few, Jolicoeur (1984) considered a small-sample test, Flury (1988) proposed an extension to more eigenvectors, Tyler (1981, 1983) robustified the test to possible (elliptical) departures from multinormality, while Schwartzman, Mascarenhas and Taylor (2008) considered extensions to the case of Gaussian random matrices. More recently, Hallin, Paindaveine and Verdebout (2010) obtained the Le Cam optimal test for the problem above. This test, ϕ_{HPV} say, rejects the null hypothesis at asymptotic level α when

$$Q_{HPV} := \frac{n}{\hat{\lambda}_1} \sum_{j=2}^p \hat{\lambda}_j^{-1} (\tilde{\theta}_j' \mathbf{S} \theta_1^0)^2 > \chi_{p-1, 1-\alpha}^2,$$

where $\tilde{\theta}_j, j = 2, \dots, p$, defined recursively through

$$(1.1) \quad \tilde{\theta}_j := \frac{(\mathbf{I}_p - \theta_1^0 \theta_1^{0'} - \sum_{k=2}^{j-1} \tilde{\theta}_k \tilde{\theta}_k') \hat{\theta}_j}{\|(\mathbf{I}_p - \theta_1^0 \theta_1^{0'} - \sum_{k=2}^{j-1} \tilde{\theta}_k \tilde{\theta}_k') \hat{\theta}_j\|}$$

(with summation over an empty collection of indices being equal to zero), result from a Gram–Schmidt orthogonalization of $\theta_1^0, \hat{\theta}_2, \dots, \hat{\theta}_p$, where $\hat{\theta}_j$ is a unit eigenvector of \mathbf{S} associated with the eigenvalue $\hat{\lambda}_j, j = 2, \dots, p$, and where we wrote \mathbf{I}_ℓ for the ℓ -dimensional identity matrix. When the eigenvalues of $\boldsymbol{\Sigma}$ are fixed and satisfy $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p$ (the minimal condition under which θ_1 is identified—up to an unimportant sign, as already mentioned), both tests above are asymptotically equivalent under the null hypothesis, hence also under sequences of contiguous alternatives, which implies that ϕ_A is also Le Cam optimal; see Hallin, Paindaveine and Verdebout (2010). The tests ϕ_A and ϕ_{HPV} can therefore be considered perfectly equivalent, at least asymptotically so.

In the present paper, we compare the asymptotic behaviors of these tests in a nonstandard asymptotic framework where eigenvalues may depend on n and where $\lambda_{n1}/\lambda_{n2}$ converges to 1 as n diverges to infinity. Such asymptotic scenarios provide *weak identifiability* since the first eigenvector θ_1 is not properly identified in the limit. To make our results as striking as possible, we will restrict to single-spiked spectra of the form $\lambda_{n1} > \lambda_{n2} = \dots = \lambda_{np}$. In other words, we will consider triangular arrays of observations \mathbf{X}_{ni} , $i = 1, \dots, n$, $n = 1, 2, \dots$, where $\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}$ form a random sample from the p -variate normal distribution with mean μ_n and covariance matrix

$$(1.2) \quad \begin{aligned} \Sigma_n &:= \sigma_n^2 (\mathbf{I}_p + r_n v \theta_1 \theta_1') \\ &= \sigma_n^2 (1 + r_n v) \theta_1 \theta_1' + \sigma_n^2 (\mathbf{I}_p - \theta_1 \theta_1'), \end{aligned}$$

where v is a positive real number, (σ_n) is a positive real sequence and (r_n) is a bounded positive real sequence (again, the multinormality assumption will be relaxed later in the paper). The eigenvalues of the covariance matrix Σ_n are then $\lambda_{n1} = \sigma_n^2 (1 + r_n v)$ (with corresponding eigenvector θ_1) and $\lambda_{n2} = \dots = \lambda_{np} = \sigma_n^2$ (with corresponding eigenspace being the orthogonal complement of θ_1 in \mathbb{R}^p). If $r_n \equiv 1$ (or more generally if r_n stays away from 0 as $n \rightarrow \infty$), then this setup is similar to the classical one where the first eigenvector θ_1 remains identified in the limit. In contrast, if $r_n = o(1)$, then the resulting weak identifiability intuitively makes the problem of testing $\mathcal{H}_0^{(n)} : \theta_1 = \theta_1^0$ against $\mathcal{H}_1^{(n)} : \theta_1 \neq \theta_1^0$ increasingly hard as n diverges to infinity.

Our results show that, while they are, as mentioned above, equivalent in the standard asymptotic scenario associated with $r_n \equiv 1$, the tests ϕ_{HPV} and ϕ_A actually exhibit very different behaviors under weak identifiability. More precisely, we show that this asymptotic equivalence survives scenarios where $r_n = o(1)$ with $\sqrt{n}r_n \rightarrow \infty$, but not scenarios where $r_n = O(1/\sqrt{n})$. Irrespective of the asymptotic scenario considered, the test ϕ_{HPV} asymptotically meets the nominal level constraint, hence may be considered robust to weak identifiability. On the contrary, in scenarios where $r_n = O(1/\sqrt{n})$, the test ϕ_A dramatically overrejects the null hypothesis. Consequently, despite the asymptotic equivalence of these tests in standard asymptotic scenarios, the test ϕ_{HPV} should be favored over ϕ_A .

Of course, this nice robustness property of ϕ_{HPV} only refers to the null asymptotic behavior of this test, and it is of interest to investigate whether or not this null robustness is obtained at the expense of power. In order to do so, we study the non-null and optimality properties of ϕ_{HPV} under suitable local alternatives. This is done by exploiting the Le Cam's asymptotic theory of statistical experiments. In every asymptotic scenario considered, we show that the corresponding sequence of experiments converges to a limiting experiment in the Le Cam sense. Interestingly, (i) the corresponding contiguity rate crucially depends on the underlying asymptotic scenario and (ii) the resulting limiting experiment is not always a Gaussian shift experiment, such as in the standard *local asymptotic normality (LAN)* setup. In all cases, however, we can derive the asymptotic non-null distribution of Q_{HPV} under contiguous alternatives by resorting to the Le Cam third lemma, and we can establish that this test enjoys excellent optimality properties.

The problem we consider in this paper is characterized by the fact that the parameter of interest (here, the first eigenvector) is unidentified when a nuisance parameter is equal to some given value (here, when the ratio of both largest eigenvalues is equal to one). Such situations have already been considered in the statistics and econometrics literatures; we refer, for example, to Dufour (1997), Pötscher (2002), Forchini and Hillier (2003), Dufour (2006), or Paindaveine and Verdebout (2017). To the best of our knowledge, however, no results have been obtained in PCA under weak identifiability. We think that, far from being of academic interest only, our results are also crucial for practitioners: they indeed provide a clear warning

that, when the underlying distribution is close to spherical (more generally, when both largest sample eigenvalues are nearly equal), the daily-practice Gaussian test ϕ_A tends to overreject the null hypothesis, hence may lead to wrong conclusions (false positives) with very high probability, whereas the test ϕ_{HPV} remains a reliable procedure in such cases. We provide an illustrative real data example that shows the practical relevance of our results.

The paper is organized as follows. In Section 2, we introduce the distributional setup and notation to be used throughout and we derive preliminary results on the asymptotic behavior of sample eigenvalues/eigenvectors. In Section 3, we show that the null asymptotic distribution of Q_{HPV} is χ_{p-1}^2 under all asymptotic scenarios, whereas that of Q_A is χ_{p-1}^2 only if $\sqrt{nr_n} \rightarrow \infty$. We also explicitly provide the null asymptotic distribution of Q_A when $r_n = O(1/\sqrt{n})$. In Section 4, we show that, in all asymptotic scenarios, the sequence of experiments considered converges to a limiting experiment. Then, this is used to study the non-null and optimality properties of ϕ_{HPV} . In Section 5, we extend our results to the more general elliptical case. Theoretical findings in Sections 3 to 5 are illustrated through Monte Carlo exercises. We treat a real data illustration in Section 6. Finally, we wrap up and shortly discuss research perspectives in Section 7. All proofs are provided in the Supplementary Material [Paindaveine, Remy and Verdebout \(2019\)](#).

2. Preliminary results. As mentioned above, we will consider throughout triangular arrays of observations \mathbf{X}_{ni} , $i = 1, \dots, n$, $n = 1, 2, \dots$, where $\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}$ form a random sample from the p -variate normal distribution with mean $\boldsymbol{\mu}_n$ and covariance matrix $\boldsymbol{\Sigma}_n = \sigma_n^2(\mathbf{I}_p + r_n v \boldsymbol{\theta}_1 \boldsymbol{\theta}_1')$, where $\boldsymbol{\theta}_1$ is a unit p -vector and σ_n, r_n and v are positive real numbers. The resulting hypothesis will be denoted as $P_{\boldsymbol{\mu}_n, \sigma_n, \boldsymbol{\theta}_1, r_n, v} = P_{\boldsymbol{\mu}_n, \sigma_n, \boldsymbol{\theta}_1, r_n, v}^{(n)}$ (the superscript (n) will be dropped in the sequel). Throughout, $\bar{\mathbf{X}}_n := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{ni}$ and $\mathbf{S}_n := \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_{ni} - \bar{\mathbf{X}}_n)(\mathbf{X}_{ni} - \bar{\mathbf{X}}_n)'$ will denote the sample average and sample covariance matrix of $\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}$. For any $j = 1, \dots, p$, the j th largest eigenvalue of \mathbf{S}_n and “the” corresponding unit eigenvector will be denoted as $\hat{\lambda}_{nj}$ and $\hat{\boldsymbol{\theta}}_{nj}$, respectively (identifiability is discussed at the end of this paragraph). With this notation, the tests $\phi_A = \phi_A^{(n)}$ and $\phi_{HPV} = \phi_{HPV}^{(n)}$ from the [Introduction](#) reject the null hypothesis at asymptotic level α when

$$\begin{aligned}
 (2.1) \quad Q_A &= Q_A^{(n)} = n(\hat{\lambda}_{n1} \boldsymbol{\theta}_1^{0'} \mathbf{S}_n^{-1} \boldsymbol{\theta}_1^0 + \hat{\lambda}_{n1}^{-1} \boldsymbol{\theta}_1^0 \mathbf{S}_n \boldsymbol{\theta}_1^0 - 2) \\
 &= \frac{n}{\hat{\lambda}_{n1}} \sum_{j=2}^p \hat{\lambda}_{nj}^{-1} (\hat{\lambda}_{n1} - \hat{\lambda}_{nj})^2 (\hat{\boldsymbol{\theta}}_{nj}' \boldsymbol{\theta}_1^0)^2 > \chi_{p-1, 1-\alpha}^2
 \end{aligned}$$

and

$$(2.2) \quad Q_{HPV} = Q_{HPV}^{(n)} = \frac{n}{\hat{\lambda}_{n1}} \sum_{j=2}^p \hat{\lambda}_{nj}^{-1} (\tilde{\boldsymbol{\theta}}_{nj}' \mathbf{S}_n \boldsymbol{\theta}_1^0)^2 > \chi_{p-1, 1-\alpha}^2,$$

respectively, where $\tilde{\boldsymbol{\theta}}_{n2}, \dots, \tilde{\boldsymbol{\theta}}_{np}$ result from the Gram–Schmidt orthogonalization in (1.1) applied to $\boldsymbol{\theta}_1^0, \hat{\boldsymbol{\theta}}_{n2}, \dots, \hat{\boldsymbol{\theta}}_{np}$. Under $P_{\boldsymbol{\mu}_n, \sigma_n, \boldsymbol{\theta}_1, r_n, v}$, the sample eigenvalue $\hat{\lambda}_{nj}$ is uniquely defined with probability one, but $\hat{\boldsymbol{\theta}}_{nj}$ is, still with probability one, defined up to a sign only. Clearly, this sign does not play any role in (2.1)–(2.2), hence will be fixed arbitrarily. At a few places below, however, this sign will need to be fixed in an appropriate way.

For obvious reasons, the asymptotic behavior of \mathbf{S}_n will play a crucial role when investigating the asymptotic properties of the tests above. To describe this behavior, we need to introduce the following notation. For an $\ell \times \ell$ matrix \mathbf{A} , denote as $\text{vec}(\mathbf{A})$ the vector obtained by stacking the columns of \mathbf{A} on top of each other. We will let $\mathbf{A}^{\otimes 2} := \mathbf{A} \otimes \mathbf{A}$, where

$\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of \mathbf{A} and \mathbf{B} . The *commutation matrix* $\mathbf{K}_{k,\ell}$, that is such that $\mathbf{K}_{k,\ell}(\text{vec } \mathbf{A}) = \text{vec}(\mathbf{A}')$ for any $k \times \ell$ matrix \mathbf{A} , satisfies $\mathbf{K}_{p,k}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{q,\ell}$ for any $k \times \ell$ matrix \mathbf{A} and $p \times q$ matrix \mathbf{B} ; see, for example, Magnus and Neudecker (2007). If \mathbf{X} is p -variate standard normal, then the covariance matrix of $\text{vec}(\mathbf{X}\mathbf{X}')$ is $\mathbf{I}_{p^2} + \mathbf{K}_p$, with $\mathbf{K}_p := \mathbf{K}_{p,p}$; the Lévy–Lindeberg central limit theorem then easily provides the following result.

LEMMA 2.1. *Fix a unit p -vector θ_1 , $v > 0$ and a bounded positive real sequence (r_n) . Then, under $\mathbf{P}_{\mathbf{0},1,\theta_1,r_n,v}$, $\sqrt{n}(\Sigma_n^{-1/2})^{\otimes 2}\text{vec}(\mathbf{S}_n - \Sigma_n)$ is asymptotically normal with mean zero and covariance matrix $\mathbf{I}_{p^2} + \mathbf{K}_p$. In particular, (i) if $r_n \equiv 1$, then $\sqrt{n} \text{vec}(\mathbf{S}_n - \Sigma_n)$ is asymptotically normal with mean zero and covariance matrix $(\mathbf{I}_{p^2} + \mathbf{K}_p)(\Sigma(v))^{\otimes 2}$, with $\Sigma(v) := \mathbf{I}_p + v\theta_1\theta_1'$; (ii) if r_n is $o(1)$, then $\sqrt{n} \text{vec}(\mathbf{S}_n - \Sigma_n)$ is asymptotically normal with mean zero and covariance matrix $\mathbf{I}_{p^2} + \mathbf{K}_p$.*

Clearly, the tests ϕ_A and ϕ_{HPV} above are invariant under translations and scale transformations, that is, respectively, under transformations of the form $(\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}) \mapsto (\mathbf{X}_{n1} + \mathbf{t}, \dots, \mathbf{X}_{nn} + \mathbf{t})$, with $\mathbf{t} \in \mathbb{R}^p$, and $(\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}) \mapsto (s\mathbf{X}_{n1}, \dots, s\mathbf{X}_{nn})$, with $s > 0$. This implies that, when investigating the behavior of these tests, we may assume without loss of generality that $\mu_n \equiv \mathbf{0}$ and $\sigma_n \equiv 1$, that is, we may restrict to hypotheses of the form $\mathbf{P}_{\theta_1,r_n,v} := \mathbf{P}_{\mathbf{0},1,\theta_1,r_n,v}$, as we already did in Lemma 2.1. We therefore restrict to such hypotheses in the rest of the paper.

The tests ϕ_A and ϕ_{HPV} are based on statistics that do not only involve the sample covariance matrix \mathbf{S}_n , but also the corresponding sample eigenvalues and eigenvectors. It is therefore no surprise that investigating the asymptotic behavior of these tests under weak identifiability will require controlling the asymptotic behaviors of sample eigenvalues and eigenvectors. For eigenvalues, we have the following result (throughout, $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_m)$ stands for the block-diagonal matrix with diagonal blocks $\mathbf{A}_1, \dots, \mathbf{A}_m$).

LEMMA 2.2. *Fix a unit p -vector θ_1 , $v > 0$ and a bounded positive real sequence (r_n) . Let $\mathbf{Z}(v)$ be a $p \times p$ random matrix such that $\text{vec}(\mathbf{Z}(v)) \sim \mathcal{N}(\mathbf{0}, (\mathbf{I}_{p^2} + \mathbf{K}_p)(\Lambda(v))^{\otimes 2})$, with $\Lambda(v) := \text{diag}(1 + v, 1, \dots, 1)$, and let $\mathbf{Z}_{22}(v)$ be the matrix obtained from $\mathbf{Z}(v)$ by deleting its first row and first column. Write $\mathbf{Z} := \mathbf{Z}(0)$ and $\mathbf{Z}_{22} := \mathbf{Z}_{22}(0)$. Then, under $\mathbf{P}_{\theta_1,r_n,v}$,*

$$\ell_n := (\sqrt{n}(\hat{\lambda}_{n1} - (1 + r_nv)), \sqrt{n}(\hat{\lambda}_{n2} - 1), \dots, \sqrt{n}(\hat{\lambda}_{np} - 1))' \xrightarrow{\mathcal{D}} \ell = (\ell_1, \dots, \ell_p)',$$

where $\xrightarrow{\mathcal{D}}$ denotes weak convergence and where ℓ is as follows:

- (i) if $r_n \equiv 1$, then ℓ_1 and $(\ell_2, \dots, \ell_p)'$ are mutually independent, ℓ_1 is normal with mean zero and variance $2(1 + v)^2$, and $\ell_2 \geq \dots \geq \ell_p$ are the eigenvalues of $\mathbf{Z}_{22}(v)$;
- (ii) if r_n is $o(1)$ with $\sqrt{nr_n} \rightarrow \infty$, then ℓ_1 and $(\ell_2, \dots, \ell_p)'$ are mutually independent, ℓ_1 is normal with mean zero and variance 2, and $\ell_2 \geq \dots \geq \ell_p$ are the eigenvalues of \mathbf{Z}_{22} ;
- (iii) if $r_n = 1/\sqrt{n}$, then ℓ_1 is the largest eigenvalue of $\mathbf{Z} - \text{diag}(0, v, \dots, v)$ and $\ell_2 \geq \dots \geq \ell_p$ are the $p - 1$ smallest eigenvalues of $\mathbf{Z} + \text{diag}(v, 0, \dots, 0)$;
- (iv) if $r_n = o(1/\sqrt{n})$, then ℓ is the vector of eigenvalues of \mathbf{Z} (in decreasing order), hence has density

$$(2.3) \quad (\ell_1, \dots, \ell_p)' \mapsto b_p \exp\left(-\frac{1}{4} \sum_{j=1}^p \ell_j^2\right) \left(\prod_{1 \leq k < j \leq p} (\ell_k - \ell_j)\right) \mathbb{I}[\ell_1 \geq \dots \geq \ell_p],$$

where b_p is a normalizing constant and where $\mathbb{I}[A]$ is the indicator function of A .

Lemma 2.2 shows that, unlike the sample covariance matrix \mathbf{S}_n , sample eigenvalues exhibit an asymptotic behavior that crucially depends on (r_n) . The important threshold, associated with $r_n = 1/\sqrt{n}$, provides sequences of hypotheses $\mathbf{P}_{\theta_1, r_n, v}$ that are contiguous to the spherical hypotheses $\mathbf{P}_{\theta_1, 0, v}$ under which the first eigenvector θ_1 is unidentified (contiguity follows, e.g., from Proposition 2.1 in Hallin and Paindaveine (2006)). Lemma 2.2 then identifies four regimes that will be present throughout our double asymptotic investigation below, namely *away from contiguity* ($r_n \equiv 1$, case (i)), *above contiguity* (r_n is $o(1)$ with $\sqrt{nr_n} \rightarrow \infty$, case (ii)), *under contiguity* ($r_n = 1/\sqrt{n}$, case (iii)), and *under strict contiguity* ($r_n = o(1/\sqrt{n})$, case (iv)).

In the high-dimensional setup where $p = p_n$ is such that $p/n \rightarrow \gamma^{-2} \in (0, 1]$, a related phase transition phenomenon has been identified in Baik, Ben Arous and Pécché (2005), in the case of complex-valued Gaussian observations. More precisely, in the single-spiked case considered in the present paper, Theorem 1.1 of Baik, Ben Arous and Pécché (2005) proves that the asymptotic behavior of $\hat{\lambda}_{n1}$ crucially depends on the ratio ρ of λ_{n1} to the common value of λ_{nj} , $j = 2, \dots, p$; there, ρ is essentially of the form $\rho = 1 + C\sqrt{p/n} (\rightarrow 1 + C\gamma^{-1})$, for some constant $C \geq 1$ whose value is showed to strongly impact the weak limit and consistency rate of $\hat{\lambda}_{n1}$. Note that, in contrast, several rates are considered for $\rho = \rho_n$ in Lemma 2.2 above, and that $\hat{\lambda}_{n1}$ exhibits the same consistency rate in each case.

While Lemmas 2.1–2.2 will be sufficient to study the asymptotic behavior of ϕ_{HPV} , the test ϕ_A , as hinted by the expression in (2.1), will further require investigating the joint asymptotic behavior of $\hat{\theta}'_{n2}\theta_1^0, \dots, \hat{\theta}'_{np}\theta_1^0$. To do so, fix arbitrary p -vectors $\theta_2, \dots, \theta_p$ such that $\Gamma := (\theta_1^0, \theta_2, \dots, \theta_p)$ is orthogonal. Let further $\hat{\Gamma}_n := (\hat{\theta}_{n1}, \dots, \hat{\theta}_{np})$, where the “signs” of $\hat{\theta}_{nj}$, $j = 1, \dots, p$, are fixed by the constraint that, with probability one, all entries in the first column of

$$(2.4) \quad \mathbf{E}_n := \hat{\Gamma}'_n \Gamma = \begin{pmatrix} E_{n,11} & \mathbf{E}_{n,12} \\ \mathbf{E}_{n,21} & \mathbf{E}_{n,22} \end{pmatrix}$$

are positive (note that all entries of \mathbf{E}_n are almost surely nonzero). With this notation, $\mathbf{E}_{n,21}$ collects the random variables $\hat{\theta}'_{n2}\theta_1^0, \dots, \hat{\theta}'_{np}\theta_1^0$ of interest above. We then have the following result.

LEMMA 2.3. *Fix a unit p -vector θ_1 , $v > 0$ and a bounded positive real sequence (r_n) . Let \mathbf{Z} be a $p \times p$ random matrix such that $\text{vec}(\mathbf{Z}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p^2} + \mathbf{K}_p)$. Let $\mathbf{E}(v) := (\mathbf{w}_1(v), \dots, \mathbf{w}_p(v))'$, where $\mathbf{w}_j(v) = (w_{j1}(v), \dots, w_{jp}(v))'$ is the unit eigenvector associated with the j th largest eigenvalue of $\mathbf{Z} + \text{diag}(v, 0, \dots, 0)$ and such that $w_{j1}(v) > 0$ almost surely. Extending the definitions to the case $v = 0$, write $\mathbf{E} := \mathbf{E}(0)$. Then, we have the following under $\mathbf{P}_{\theta_1, r_n, v}$:*

- (i) *if $r_n \equiv 1$, then $E_{n,11} = 1 + o_P(1)$, $\mathbf{E}_{n,22}\mathbf{E}'_{n,22} = \mathbf{I}_{p-1} + o_P(1)$, $\sqrt{n}\mathbf{E}_{n,21} = O_P(1)$, and both $\sqrt{n}\mathbf{E}'_{n,22}\mathbf{E}_{n,21}$ and $\sqrt{n}\mathbf{E}'_{n,12}$ are asymptotically normal with mean zero and covariance matrix $v^{-2}(1+v)\mathbf{I}_{p-1}$;*
- (ii) *if r_n is $o(1)$ with $\sqrt{nr_n} \rightarrow \infty$, then $E_{n,11} = 1 + o_P(1)$, $\mathbf{E}_{n,22}\mathbf{E}'_{n,22} = \mathbf{I}_{p-1} + o_P(1)$, $\sqrt{nr_n}\mathbf{E}_{n,21} = O_P(1)$, and both $\sqrt{nr_n}\mathbf{E}'_{n,22}\mathbf{E}_{n,21}$ and $\sqrt{nr_n}\mathbf{E}'_{n,12}$ are asymptotically normal with mean zero and covariance matrix $v^{-2}\mathbf{I}_{p-1}$;*
- (iii) *if $r_n = 1/\sqrt{n}$, then \mathbf{E}_n converges weakly to $\mathbf{E}(v)$;*
- (iv) *if $r_n = o(1/\sqrt{n})$, then \mathbf{E}_n converges weakly to \mathbf{E} .*

This result shows that the asymptotic behavior of $\mathbf{E}_{n,21}$, which, as mentioned above, is the only part of \mathbf{E}_n involved in the Anderson test statistic Q_A , depends on the regimes identified in Lemma 2.2. (i) Away from contiguity, $\mathbf{E}_{n,21}$ converges to the zero vector in probability at

the standard root- n rate. (ii) Above contiguity, $\mathbf{E}_{n,21}$ is still $op(1)$, but the rate of convergence is slower. (iii) Under contiguity, consistency is lost and $\mathbf{E}_{n,21}$ converges weakly to a distribution that still depends on v . (iv) Under strict contiguity, on the contrary, the asymptotic distribution of \mathbf{E}_n does not depend on v and inspection of the proof of Lemma 2.3 reveals that this asymptotic distribution is the same as the one we would obtain for $v = 0$. In other words, the asymptotic distribution of \mathbf{E}_n is then the same as in the spherical Gaussian case, so that, up to the fact that \mathbf{E} has almost surely positive entries in its first column (a constraint inherited from the corresponding one on \mathbf{E}_n), this asymptotic distribution is the invariant Haar distribution on the group of $p \times p$ orthogonal matrices; see Anderson (1963), page 126.

3. Null results. In this section, we will study the null asymptotic behaviors of ϕ_A and ϕ_{HPV} under weak identifiability, that is, we do so under the sequences of (null) hypotheses $\mathbf{P}_{\theta_1^0, r_n, v}^0$ introduced in the previous section. Before doing so theoretically, we consider the following Monte Carlo exercise. For any $\ell = 0, 1, \dots, 5$, we generated $M = 10,000$ mutually independent random samples $\mathbf{X}_i^{(\ell)}$, $i = 1, \dots, n$, from the ($p = 10$)-variate normal distribution with mean zero and covariance matrix

$$(3.1) \quad \Sigma_n^{(\ell)} := \mathbf{I}_p + n^{-\ell/6} \theta_1^0 \theta_1^{0'}$$

where θ_1^0 is the first vector of the canonical basis of \mathbb{R}^p . For each sample, we performed the tests ϕ_{HPV} and ϕ_A for $\mathcal{H}_0^{(n)} : \theta_1 = \theta_1^0$ at nominal level 5%. The value of ℓ allows to consider the various regimes above, namely (i) away from contiguity ($\ell = 0$), (ii) beyond contiguity ($\ell = 1, 2$), (iii) under contiguity ($\ell = 3$), and (iv) under strict contiguity ($\ell = 4, 5$). Increasing values of ℓ therefore provide harder and harder inference problems. Figure 1, that reports the resulting rejection frequencies for $n = 200$ and $n = 500,000$, suggests that ϕ_{HPV} asymptotically shows the target Type 1 risk in all regimes, hence is *validity-robust* to weak identifiability. In sharp contrast, ϕ_A seems to exhibit the right asymptotic null size in regimes (i)–(ii) only, as it dramatically overrejects the null hypothesis (even asymptotically) in regimes (iii)–(iv).

We now turn to the theoretical investigation of the null asymptotic behaviors of ϕ_A and ϕ_{HPV} under weak identifiability. Obviously, this will heavily rely on Lemmas 2.2–2.3. For ϕ_{HPV} , we have the following result.

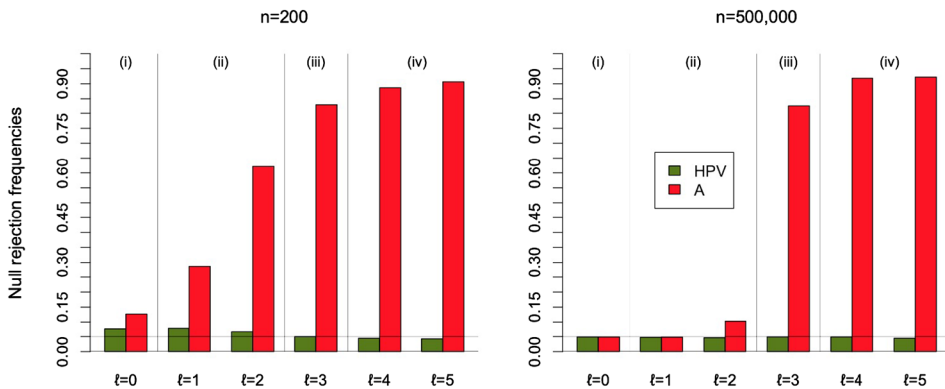


FIG. 1. Empirical rejection frequencies, under the null hypothesis, of the tests ϕ_{HPV} and ϕ_A performed at nominal level 5%. Results are based on $M = 10,000$ independent ten-dimensional Gaussian random samples of size $n = 200$ (left) and size $n = 500,000$ (right). Increasing values of ℓ bring the underlying spiked covariance matrix closer and closer to a multiple of the identity matrix; see Section 3 for details. The link between the values of ℓ and the asymptotic regimes (i)–(iv) from Section 2 is provided in each barplot.

THEOREM 3.1. Fix a unit p -vector θ_1^0 , $v > 0$ and a bounded positive real sequence (r_n) . Then, under $P_{\theta_1^0, r_n, v}$,

$$Q_{\text{HPV}} \xrightarrow{\mathcal{D}} \chi_{p-1}^2,$$

so that, in all regimes (i)–(iv) from the previous section, the test ϕ_{HPV} has asymptotic size α under the null hypothesis.

This result confirms that the test ϕ_{HPV} is validity-robust to weak identifiability in the sense that it will asymptotically meet the nominal level constraint in scenarios that are arbitrarily close to the spherical case. As hinted by the above Monte Carlo exercise, the situation is more complex for the Anderson test ϕ_A . We have the following result.

THEOREM 3.2. Fix a unit p -vector θ_1^0 , $v > 0$ and a bounded positive real sequence (r_n) . Let \mathbf{Z} be a $p \times p$ random matrix such that $\text{vec}(\mathbf{Z}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p^2} + \mathbf{K}_p)$. Then, we have the following under $P_{\theta_1^0, r_n, v}$:

(i)–(ii) if $r_n \equiv 1$ or if r_n is $o(1)$ with $\sqrt{nr_n} \rightarrow \infty$, then

$$Q_A \xrightarrow{\mathcal{D}} \chi_{p-1}^2,$$

so that the test ϕ_A has asymptotic size α under the null hypothesis;

(iii) if $r_n = 1/\sqrt{n}$, then

$$Q_A \xrightarrow{\mathcal{D}} \sum_{j=2}^p (\ell_1(v) - \ell_j(v))^2 (w_{j1}(v))^2,$$

where $\ell_1(v) \geq \dots \geq \ell_p(v)$ are the eigenvalues of $\mathbf{Z} + \text{diag}(v, 0, \dots, 0)$ and $\mathbf{w}_j(v) = (w_{j1}(v), \dots, w_{jp}(v))'$ is an arbitrary unit eigenvector associated with $\ell_j(v)$ (with probability one, the only freedom in the choice of $\mathbf{w}_j(v)$ is in the sign of $w_{j1}(v)$, that is clearly irrelevant here);

(iv) if $r_n = o(1/\sqrt{n})$, then

$$Q_A \xrightarrow{\mathcal{D}} \sum_{j=2}^p (\ell_1 - \ell_j)^2 w_{j1}^2,$$

where $\ell_1 \geq \dots \geq \ell_p$ are the eigenvalues of \mathbf{Z} and $\mathbf{w}_j = (w_{j1}, \dots, w_{jp})'$ is an arbitrary unit eigenvector associated with ℓ_j .

This result ensures that the Anderson test asymptotically meets the nominal level constraint in regimes (i)–(ii). To see whether or not this extends to regimes (iii)–(iv), we need to investigate the asymptotic distributions in Theorem 3.2(iii)–(iv). We consider first the asymptotic distribution of Q_A under $P_{\theta_1^0, 1/\sqrt{n}, v}$, that is, in the contiguity regime (iii). To do so, we generated, for various dimensions p and for each $v = 8(\ell - 1)/19$, $\ell = 1, \dots, 20$, in a regular grid of 20 v -values in $[0, 8]$, a collection of $M = 10^6$ independent values $Z_1(v), \dots, Z_M(v)$ from the asymptotic distribution in Theorem 3.2(iii). For each p and v , we then recorded

$$(3.2) \quad r_{p, 0.95}^{(\text{iii})}(v) := \frac{1}{M} \sum_{m=1}^M \mathbb{I}[Z_m(v) > \chi_{p-1, 0.95}^2],$$

which is an excellent approximation of the asymptotic null size of the 5%-level Anderson test under $P_{\theta_1^0, 1/\sqrt{n}, v}$ (a 99%-confidence interval for the true asymptotic size has a length smaller

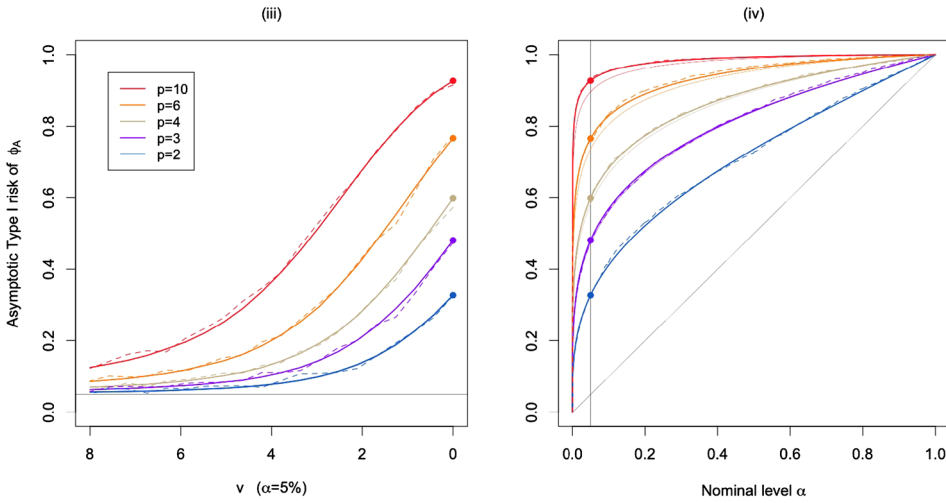


FIG. 2. (Left) Plots, for various dimensions p , of the approximate asymptotic Type I risk $r_{p,0.95}^{(iii)}(v)$ (see (3.2)) of the 5%-level Anderson test for $\mathcal{H}_0 : \theta_1 = \theta_1^0$ under $\mathbb{P}_{\theta_1^0, r_n, v}$, with $r_n = 1/\sqrt{n}$ (regime (iii)). The dashed curves report the corresponding rejection frequencies for sample size $n = 10,000$ (a regular grid of 20 v -values in $[0, 8]$ was considered and rejection frequencies were computed from 2,500 independent replications in each case). (Right) Plots, for the same dimensions p , of the approximate asymptotic Type I risk $r_{p,\alpha}^{(iv)}$ (see (3.3)) of the level- α Anderson test for $\mathcal{H}_0 : \theta_1 = \theta_1^0$ under $\mathbb{P}_{\theta_1^0, r_n, v}$, with $r_n = o(1/\sqrt{n})$ (regime (iv)). The dashed curves report the corresponding rejection frequencies computed from 2500 independent standard normal samples of size $n = 10,000$. The thin curves represent what the asymptotic Type I risk of the level- α Anderson test would be if the null asymptotic distribution of Q_A in regime (iv) would be $4\chi_{p-1}^2$; see the discussion below Corollary 3.1.

than 0.0026). The left panel of Figure 2 plots $r_{p,0.95}^{(iii)}(v)$ as a function of v for several dimensions p . Clearly, the Anderson test is, irrespective of p and v , asymptotically overrejecting the null hypothesis. The asymptotic Type I risk increases with p and decreases with v (letting v go to infinity essentially provides regime (ii), which explains that the Type I risk then converges to the nominal level). In the right panel of Figure 2, we generated, still for various values of p , $M = 10^6$ independent values Z_1, \dots, Z_M from the asymptotic distribution in Theorem 3.2(iv) and plotted the function mapping α to

$$(3.3) \quad r_{p,\alpha}^{(iv)} := \frac{1}{M} \sum_{m=1}^M \mathbb{I}[Z_m > \chi_{p-1, 1-\alpha}^2],$$

which accurately approximates the asymptotic Type I risk of the level- α Anderson test in regime (iv). Irrespective of α , the Anderson test is still overrejecting the null hypothesis asymptotically and the asymptotic Type I risk increases with p . Overrejection is dramatic: for instance, in dimension 10, the asymptotic Type I risk of the 5%-level Anderson test exceeds 92%. Empirical rejection frequencies of the Anderson test, which are also showed in Figure 2, clearly support the asymptotic results in Theorem 3.2(iii)–(iv).

The asymptotic distributions in Theorem 3.2(iii)–(iv) are explicitly described yet are quite complicated. Remarkably, for regime (iv), the asymptotic distribution is a classical one in the bivariate case $p = 2$. More precisely, we have the following result.

COROLLARY 3.1. Fix $p = 2$, a unit p -vector θ_1^0 , $v > 0$ and a positive real sequence (r_n) such that $\sqrt{nr_n} \rightarrow 0$. Then, under $\mathbb{P}_{\theta_1^0, r_n, v}$,

$$Q_A \xrightarrow{\mathcal{D}} 4\chi_1^2,$$

so that, irrespective of $\alpha \in (0, 1)$, the asymptotic size of ϕ_A under the null hypothesis is strictly larger than α .

This result shows in a striking way the impact weak identifiability may have, in the bivariate case, on the null asymptotic distribution of the Anderson test statistic Q_A : away from contiguity and beyond contiguity (regimes (i)–(ii)), Q_A is asymptotically χ_1^2 under the null hypothesis (Theorem 3.2), whereas under strict contiguity (regime (iv)), this statistic is asymptotically $4\chi_1^2$ under the null hypothesis. The result also allows us to quantify, for any nominal level α , how much the bivariate Anderson test will asymptotically overreject the null hypothesis in regime (iv). More precisely, the Type 1 risk, in this regime, of the level- α bivariate Anderson test converges to $P[4Z^2 > \chi_{1,1-\alpha}^2]$, where Z is standard normal. For $\alpha = 0.1\%$, 1% and 5% , this provides in regime (iv) an asymptotic Type 1 risk of about 10% , 19.8% and 32.7% , respectively (which exceeds the nominal level by about a factor 100, 20 and 6.5, respectively)! In dimensions $p \geq 3$, the null asymptotic distribution of Q_A in regime (iv), as showed in the right panel of Figure 2, is very close to $4\chi_{p-1}^2$, particularly so for $p = 3$. Yet the distribution is not $4\chi_{p-1}^2$. For instance, in dimension $p = 3$, it can be showed that the null asymptotic distribution of Q_A in regime (iv) has mean $49/6$, whereas the distribution $4\chi_2^2$ has mean $8 = 48/6$ (also, computing the variance of the null asymptotic distribution of Q_A shows that this distribution is not of the form $\lambda\chi_2^2$ for any $\lambda > 0$).

We close this section with a last simulation illustrating Theorem 3.2 and its consequences. To do so, we generated, for any $\ell = 0, 1, \dots, 5$, a collection of $M = 10,000$ mutually independent random samples of size $n = 500,000$ from the bivariate normal distribution with mean zero and covariance matrix $\Sigma_n^{(\ell)} := \mathbf{I}_2 + n^{-\ell/6}\theta_1^0\theta_1^{0'}$, with $\theta_1^0 = (1, 0)'$. This is therefore essentially the bivariate version of the ten-dimensional numerical exercise leading to Figure 1. For each value of ℓ , Figure 3 provides histograms of the resulting M values of the Anderson test statistic Q_A , along with plots of the densities of the χ_1^2 and $4\chi_1^2$ distributions, that is, densities of the null asymptotic distribution of Q_A in regimes (i)–(ii) and in regime (iv), respectively. In these three regimes, the histograms are perfectly fitted by the corresponding density. The figure also provides the empirical Type 1 risks of the level- α Anderson test for $\alpha = 0.1\%$, 1% and 5% . Clearly, these Type 1 risks are close to the theoretical asymptotic ones both in regimes (i)–(ii) (namely, α) and in regime (iv) (namely, the Type 1 risks provided in the previous paragraph).

4. Non-null and optimality results. The previous section shows that, unlike ϕ_A , the test ϕ_{HPV} is validity-robust to weak identifiability. However, the trivial level- α test, that randomly rejects the null hypothesis with probability α , of course enjoys the same robustness property. This motivates investigating whether or not the validity-robustness of ϕ_{HPV} is obtained at the expense of power. In this section, we therefore study the asymptotic non-null behavior of ϕ_{HPV} and show that this test actually still enjoys strong optimality properties under weak identifiability.

Throughout, optimality is to be understood in the Le Cam sense. In the present hypothesis testing context, Le Cam optimality requires studying local log-likelihood ratios of the form

$$(4.1) \quad \Lambda_n := \log \frac{dP_{\theta_1^0 + v_n \tau_n, r_n, v}}{dP_{\theta_1^0, r_n, v}},$$

where the bounded sequence (τ_n) in \mathbb{R}^p and the positive real sequence (v_n) are such that, for any n , $\theta_1^0 + v_n \tau_n$ is a unit p -vector, hence, is an admissible perturbation of θ_1^0 . This imposes that (τ_n) satisfies

$$(4.2) \quad \theta_1^{0'} \tau_n = -\frac{v_n}{2} \|\tau_n\|^2$$

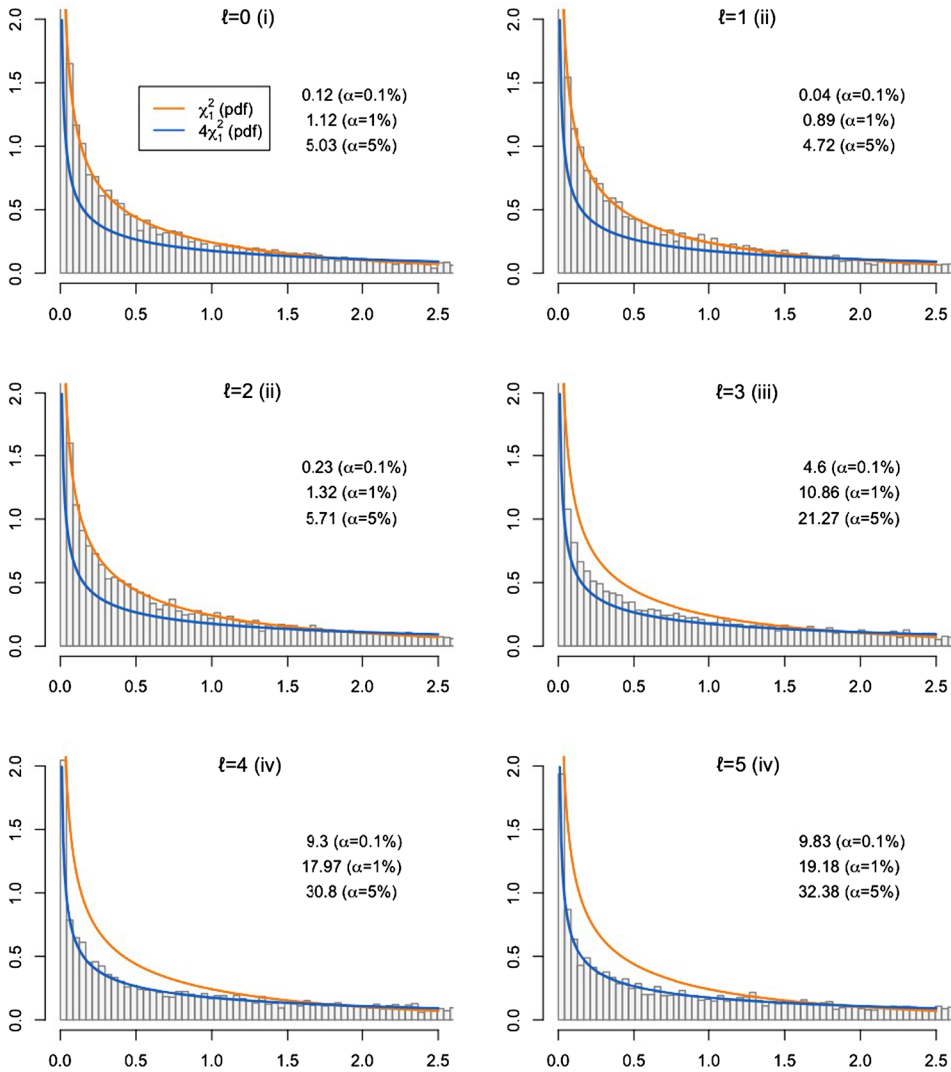


FIG. 3. For each $\ell = 0, 1, \dots, 5$, histograms of values of Q_A from $M = 10,000$ independent (null) Gaussian random samples of size $n = 500,000$ and dimension $p = 2$. Increasing values of ℓ bring the underlying spiked covariance matrix closer and closer to a multiple of the identity matrix; see Section 3 for details. The links between the values of ℓ and the asymptotic regimes (i)–(iv) from Section 2 are provided in each case. For any value of ℓ , the density of the null asymptotic distribution of Q_A in regimes (i)–(ii) (resp., in regime (iv)) is plotted in orange (resp., in blue) and the empirical Type 1 risk of the level- α Anderson test is provided for $\alpha = 0.1\%$, 1% and 5% .

for any n . The following result then describes, in each of the four regimes (i)–(iv) considered in the previous sections, the asymptotic behavior of the log-likelihood ratio Δ_n .

THEOREM 4.1. Fix a unit p -vector θ_1^0 , $v > 0$ and a bounded positive real sequence (r_n) . Then, we have the following under $P_{\theta_1^0, r_n, v}$:

(i) if $r_n \equiv 1$, then, with $v_n = 1/\sqrt{n}$,

$$\Delta_n = \frac{v}{1+v} \sqrt{n} (\mathbf{I}_p - \theta_1^0 \theta_1^{0'}) (\mathbf{S}_n - \Sigma_n) \theta_1^0 \quad \text{and} \quad \Gamma = \frac{v^2}{1+v} (\mathbf{I}_p - \theta_1^0 \theta_1^{0'}),$$

we have that $\Delta_n = \tau_n' \Delta_n - \frac{1}{2} \tau_n' \Gamma \tau_n + o_P(1)$ and that Δ_n is asymptotically normal with mean zero and covariance matrix Γ ;

(ii) if r_n is $o(1)$ with $\sqrt{nr_n} \rightarrow \infty$, then, with $v_n = 1/(\sqrt{nr_n})$,

$$\Delta_n = v\sqrt{n}(\mathbf{I}_p - \boldsymbol{\theta}_1^0\boldsymbol{\theta}_1^{0'}) (\mathbf{S}_n - \boldsymbol{\Sigma}_n)\boldsymbol{\theta}_1^0 \quad \text{and} \quad \boldsymbol{\Gamma} = v^2(\mathbf{I}_p - \boldsymbol{\theta}_1^0\boldsymbol{\theta}_1^{0'}),$$

we similarly have that $\Delta_n = \boldsymbol{\tau}'_n \Delta_n - \frac{1}{2}\boldsymbol{\tau}'_n \boldsymbol{\Gamma} \boldsymbol{\tau}_n + o_P(1)$ and that Δ_n is still asymptotically normal with mean zero and covariance matrix $\boldsymbol{\Gamma}$;

(iii) if $r_n = 1/\sqrt{n}$, then, letting $v_n \equiv 1$,

$$(4.3) \quad \begin{aligned} \Delta_n &= \boldsymbol{\tau}'_n \left[v\sqrt{n}(\mathbf{S}_n - \boldsymbol{\Sigma}_n) \left(\boldsymbol{\theta}_1^0 + \frac{1}{2}\boldsymbol{\tau}_n \right) \right] \\ &\quad - \frac{v^2}{2} \|\boldsymbol{\tau}_n\|^2 + \frac{v^2}{8} \|\boldsymbol{\tau}_n\|^4 + o_P(1), \end{aligned}$$

where, if $(\boldsymbol{\tau}_n) \rightarrow \boldsymbol{\tau}$, then $\boldsymbol{\tau}'_n \sqrt{n}(\mathbf{S}_n - \boldsymbol{\Sigma}_n)(\boldsymbol{\theta}_1^0 + \frac{1}{2}\boldsymbol{\tau}_n)$ is asymptotically normal with mean zero and variance $\|\boldsymbol{\tau}\|^2 - \frac{1}{4}\|\boldsymbol{\tau}\|^4$;

(iv) if $r_n = o(1/\sqrt{n})$, then, even with $v_n \equiv 1$, we have that Δ_n is $o_P(1)$.

This result shows that, for any fixed $v > 0$ and for any fixed sequence (r_n) associated with regime (i) (away from contiguity) or regime (ii) (beyond contiguity), the sequence of models $\{\mathbf{P}_{\boldsymbol{\theta}_1, r_n, v} : \boldsymbol{\theta}_1 \in \mathcal{S}^{p-1}\}$ is LAN (locally asymptotically normal), with central sequence

$$(4.4) \quad \Delta_{n,\delta} := \frac{\sqrt{nv}}{1 + \delta v} (\mathbf{I}_p - \boldsymbol{\theta}_1^0\boldsymbol{\theta}_1^{0'}) (\mathbf{S}_n - \boldsymbol{\Sigma}_n)\boldsymbol{\theta}_1^0$$

and Fisher information matrix

$$(4.5) \quad \boldsymbol{\Gamma}_\delta := \frac{v^2}{1 + \delta v} (\mathbf{I}_p - \boldsymbol{\theta}_1^0\boldsymbol{\theta}_1^{0'}),$$

where $\delta := 1$ if regime (i) is considered and $\delta := 0$ otherwise. Denoting as \mathbf{A}^- the Moore–Penrose inverse of \mathbf{A} , it follows that the locally asymptotically maximin test for $\mathcal{H}_0^{(n)} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ against $\mathcal{H}_1^{(n)} : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0$ rejects the null hypothesis at asymptotic level α when

$$(4.6) \quad Q_\delta = \Delta'_{n,\delta} \boldsymbol{\Gamma}_\delta^- \Delta_{n,\delta} = \frac{n}{1 + \delta v} \boldsymbol{\theta}_1^{0'} \mathbf{S}_n (\mathbf{I}_p - \boldsymbol{\theta}_1^0\boldsymbol{\theta}_1^{0'}) \mathbf{S}_n \boldsymbol{\theta}_1^0 > \chi_{p-1, 1-\alpha}^2.$$

In view of (S.2.1) in the proof of Theorem 3.1, we have that $Q_{\text{HPV}} = Q_\delta + o_P(1)$ under $\mathbf{P}_{\boldsymbol{\theta}_1^0, r_n, v}$, for any $v > 0$ and any bounded positive sequence (r_n) , hence also, from contiguity, under local alternatives of the form $\mathbf{P}_{\boldsymbol{\theta}_1^0 + \boldsymbol{\tau}_n/(\sqrt{nr_n}), r_n, v}$. We may therefore conclude that, away from contiguity and beyond contiguity, the test ϕ_{HPV} is Le Cam optimal for the problem at hand. We have the following result.

THEOREM 4.2. *Fix a unit p -vector $\boldsymbol{\theta}_1^0$, $v > 0$ and a positive real sequence (r_n) satisfying (i) $r_n \equiv 1$ or (ii) $r_n = o(1)$ with $\sqrt{nr_n} \rightarrow \infty$. Then, the test ϕ_{HPV} is locally asymptotically maximin at level α when testing $\mathcal{H}_0^{(n)} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0$ against $\mathcal{H}_1^{(n)} : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_1^0$. Moreover, under $\mathbf{P}_{\boldsymbol{\theta}_1^0 + \boldsymbol{\tau}_n/(\sqrt{nr_n}), r_n, v}$, with $(\boldsymbol{\tau}_n) \rightarrow \boldsymbol{\tau}$, the statistic Q_{HPV} is asymptotically noncentral chi-square with $p - 1$ degrees of freedom and with noncentrality parameter $(v^2/(1 + \delta v))\|\boldsymbol{\tau}\|^2$.*

Under strict contiguity (Theorem 4.1(iv)), no asymptotic level- α test can show nontrivial asymptotic powers against the most severe alternatives of the form $\boldsymbol{\theta}_1^0 + \boldsymbol{\tau}$. Therefore, the test ϕ_{HPV} is also optimal in regime (iv), even though this optimality is degenerate since the trivial level- α test is also optimal in this regime. We then turn to Theorem 4.1(iii), where the situation is much less standard, as the sequence of experiments $\{\mathbf{P}_{\boldsymbol{\theta}_1, 1/\sqrt{n}, v} : \boldsymbol{\theta}_1 \in \mathcal{S}^{p-1}\}$ there

is neither LAN, nor LAMN (locally asymptotically mixed normal), nor LAQ (locally asymptotically quadratic); see Jeganathan (1995) and Roussas and Bhattacharya (2011). While, to the best of our knowledge, the form of optimal tests in such nonstandard limiting experiments remains unknown, we will still be able below to draw conclusions about optimality for small values of τ . Before doing so, note that, by using the Le Cam first lemma (see, e.g., Lemma 6.4 in van der Vaart (1998)), Theorem 4.1(iii) readily entails that, for any $v > 0$, the sequences of hypotheses $P_{\theta_1^0, 1/\sqrt{n}, v}$ and $P_{\theta_1^0 + \tau_n, 1/\sqrt{n}, v}$, with $(\tau_n) \rightarrow \tau$, are mutually contiguous. Consequently, the asymptotic non-null distribution of the test statistic Q_{HPV} under contiguous alternatives may still be obtained from the Le Cam third lemma. We have the following result.

THEOREM 4.3. *Fix a unit p -vector θ_1^0 , $v > 0$ and a positive real sequence (r_n) satisfying (iii) $r_n = 1/\sqrt{n}$ or (iv) $r_n = o(1/\sqrt{n})$. Then, in case (iii), the test statistic Q_{HPV} , under $P_{\theta_1^0 + \tau_n, r_n, v}$, with $(\tau_n) \rightarrow \tau$, is asymptotically noncentral chi-square with $p - 1$ degrees of freedom and with noncentrality parameter*

$$(4.7) \quad \frac{v^2}{16} \|\tau\|^2 (4 - \|\tau\|^2) (2 - \|\tau\|^2)^2,$$

so that the test ϕ_{HPV} is rate-consistent. In case (iv), this test is locally asymptotically maximin at level α when testing $\mathcal{H}_0^{(n)} : \theta_1 = \theta_1^0$ against $\mathcal{H}_1^{(n)} : \theta_1 \neq \theta_1^0$, but trivially so since its asymptotic power against any sequence of alternatives $P_{\theta_1^0 + \tau_n, r_n, v}$, with $(\tau_n) \rightarrow \tau$, is then equal to α .

Figure 4 plots the noncentrality parameter in (4.7) as a function of $\tau \in [0, \sqrt{2}]$ (since θ_1 is defined up to a sign, one may restrict to alternatives $\theta_1^0 + \tau$ in the hemisphere centered at θ_1^0), as well as the resulting asymptotic powers of the test ϕ_{HPV} in dimensions $p = 2, 3$. This test shows no asymptotic power when $\|\tau\| = \sqrt{2}$ (that is, when θ_1 is orthogonal to θ_1^0), hence clearly does not enjoy *global-in- τ* optimality properties in regime (iii). As we now explain, however, ϕ_{HPV} exhibits excellent *local-in- τ* optimality properties in this regime. In order to see this, note that decomposing \mathbf{I}_p into $(\mathbf{I}_p - \theta_1^0 \theta_1^{0'}) + \theta_1^0 \theta_1^{0'}$ and using repeatedly (4.2) allows

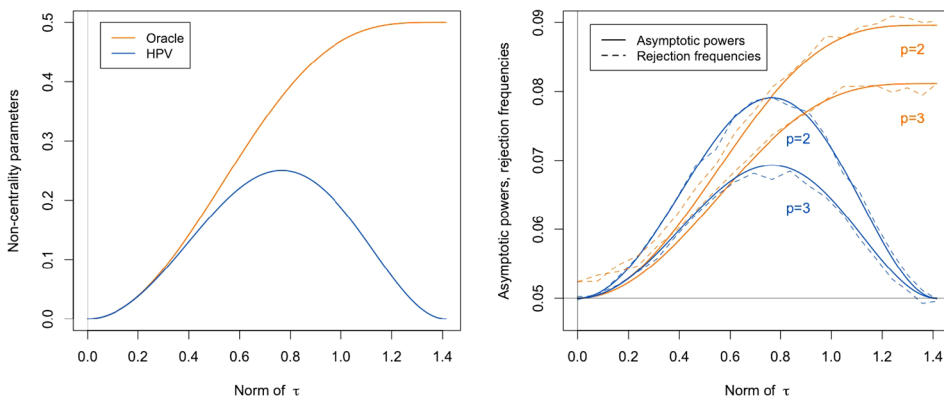


FIG. 4. (Left) Noncentrality parameters (4.7) and (4.10), as a function of $\|\tau\| \in [0, \sqrt{2}]$, in the asymptotic noncentral chi-square distributions of the test statistics of ϕ_{HPV} and ϕ_{Oracle} , respectively, under alternatives of the form $P_{\theta_1^0 + \tau, 1/\sqrt{n}, 1}$. (Right) The corresponding asymptotic power curves in dimensions $p = 2$ and $p = 3$, as well as the empirical power curves resulting from the Monte Carlo exercise described at the end of Section 4.

us to rewrite (4.3) as

$$(4.8) \quad \begin{aligned} \Lambda_n &= \boldsymbol{\tau}'_n \mathbf{\Delta}_{n,0} - \frac{1}{2} \boldsymbol{\tau}'_n \mathbf{\Gamma}_0 \boldsymbol{\tau}_n \\ &\quad - \frac{\sqrt{nv}}{2} \|\boldsymbol{\tau}_n\|^2 \boldsymbol{\theta}_1^{0'} (\mathbf{S}_n - \boldsymbol{\Sigma}_n) \boldsymbol{\theta}_1^0 + \frac{\sqrt{nv}}{2} \boldsymbol{\tau}'_n (\mathbf{S}_n - \boldsymbol{\Sigma}_n) \boldsymbol{\tau}_n + o_P(1), \end{aligned}$$

where $\mathbf{\Delta}_{n,0}$ and $\mathbf{\Gamma}_0$ were defined in (4.4) and (4.5), respectively. For small perturbations $\boldsymbol{\tau}_n$, the righthand side of (4.8), after neglecting the second-order random terms in $\boldsymbol{\tau}_n$, becomes

$$\boldsymbol{\tau}'_n \mathbf{\Delta}_{n,0} - \frac{1}{2} \boldsymbol{\tau}'_n \mathbf{\Gamma}_0 \boldsymbol{\tau}_n + o_P(1),$$

so that the sequence of experiments is then LAN again, with central sequence $\mathbf{\Delta}_{n,0}$ and Fisher information matrix $\mathbf{\Gamma}_0$. This implies that the test in (4.6) and (in view of the asymptotic equivalence stated below (4.6)) the test ϕ_{HPV} are locally(-in- $\boldsymbol{\tau}$) asymptotically maximin.

Now, if the objective is to construct a test that will perform well also for large perturbations $\boldsymbol{\tau}_n$ in regime (iii), it may be tempting to consider as a test statistic the linear-in- $\boldsymbol{\tau}$ part of the random term in (4.3), namely $\tilde{\mathbf{\Delta}}_n := v\sqrt{n}(\mathbf{S}_n - \boldsymbol{\Sigma}_n)\boldsymbol{\theta}_1^0$. Since $\tilde{\mathbf{\Delta}}_n$, under $P_{\boldsymbol{\theta}_1^0, 1/\sqrt{n}, v}$, is asymptotically normal with mean zero and covariance $\tilde{\mathbf{\Gamma}} := v^2(\mathbf{I}_p + \boldsymbol{\theta}_1^0 \boldsymbol{\theta}_1^{0'})$, the resulting test, ϕ_{oracle} say, rejects the null hypothesis at asymptotic level α when

$$(4.9) \quad \begin{aligned} \tilde{Q} &:= \tilde{\mathbf{\Delta}}_n \tilde{\mathbf{\Gamma}}^{-1} \tilde{\mathbf{\Delta}}_n = n \boldsymbol{\theta}_1^{0'} (\mathbf{S}_n - \boldsymbol{\Sigma}_n) (\mathbf{I}_p + \boldsymbol{\theta}_1^0 \boldsymbol{\theta}_1^{0'})^{-1} (\mathbf{S}_n - \boldsymbol{\Sigma}_n) \boldsymbol{\theta}_1^0 \\ &= n \left(\mathbf{S}_n \boldsymbol{\theta}_1^0 - \left(1 + \frac{v}{\sqrt{n}}\right) \boldsymbol{\theta}_1^0 \right)' \left(\mathbf{I}_p - \frac{1}{2} \boldsymbol{\theta}_1^0 \boldsymbol{\theta}_1^{0'} \right) \left(\mathbf{S}_n \boldsymbol{\theta}_1^0 - \left(1 + \frac{v}{\sqrt{n}}\right) \boldsymbol{\theta}_1^0 \right) \\ &> \chi_{p, 1-\alpha}^2; \end{aligned}$$

the terminology ‘‘oracle’’ stresses that this test requires knowing the true value of v . The Le Cam third lemma entails that $\tilde{\mathbf{\Delta}}_n$, under $P_{\boldsymbol{\theta}_1^0 + \boldsymbol{\tau}_n, 1/\sqrt{n}, v}$, with $(\boldsymbol{\tau}_n) \rightarrow \boldsymbol{\tau}$, is asymptotically normal with mean $v^2(1 - \frac{1}{2}\|\boldsymbol{\tau}\|^2)\boldsymbol{\tau} - (v^2/2)\|\boldsymbol{\tau}\|^2 \boldsymbol{\theta}_1^0$ and covariance matrix $\tilde{\mathbf{\Gamma}}$. Therefore, under the same sequence of hypotheses, \tilde{Q} is asymptotically noncentral chi-square with p degrees of freedom and with noncentrality parameter

$$(4.10) \quad \frac{v^2}{16} \|\boldsymbol{\tau}\|^2 (4 - \|\boldsymbol{\tau}\|^2) \left(4 - 2\|\boldsymbol{\tau}\|^2 + \frac{1}{2}\|\boldsymbol{\tau}\|^4 \right).$$

Note that the difference between this noncentrality parameter and the one in (4.7) is $O(\|\boldsymbol{\tau}\|^4)$ as $\|\boldsymbol{\tau}\|$ goes to zero. Since ϕ_{HPV} is based on a smaller number of degrees of freedom ($p - 1$, versus p for the oracle test), it will therefore exhibit larger asymptotic powers than the oracle test for small values of $\boldsymbol{\tau}$, which reflects the aforementioned local-in- $\boldsymbol{\tau}$ optimality of ϕ_{HPV} .

Figure 4 also plots the noncentrality parameter in (4.10) as a function of $\|\boldsymbol{\tau}\|$, as well as the asymptotic powers of the oracle test in dimensions $p = 2, 3$. As predicted above, ϕ_{HPV} dominates ϕ_{oracle} for small values of $\|\boldsymbol{\tau}\|$, that is, for small perturbations. The opposite happens for large values of the perturbation and it is seen that ϕ_{oracle} overall is quite efficient. It is important to recall, however, that this test cannot be used in practice since it requires knowing the value of v . The figure further reports the results of a Monte Carlo exercise we conducted to check correctness of the highly nonstandard asymptotic results obtained in the present regime (iii). In this simulation, we generated $M = 200,000$ mutually independent random samples $\mathbf{X}_i^{(k)}$, $i = 1, \dots, n = 10,000$, $k = 0, 1, \dots, 20$, of p -variate ($p = 2, 3$) Gaussian random vectors with mean zero and covariance matrix

$$(4.11) \quad \boldsymbol{\Sigma}_n^{(k)} := \mathbf{I}_p + n^{-1/2}(\boldsymbol{\theta}_1^0 + \boldsymbol{\tau}_k)(\boldsymbol{\theta}_1^0 + \boldsymbol{\tau}_k)',$$

where $\theta_1^0 = (1, 0, \dots, 0)' \in \mathbb{R}^p$ and $\theta_1^0 + \tau_k = (\cos(k\pi/40), \sin(k\pi/40), 0, \dots, 0)' \in \mathbb{R}^p$. In each replication, we performed the tests ϕ_{HPV} and ϕ_{oracle} for $\mathcal{H}_0^{(n)} : \theta_1 = \theta_1^0$ at nominal level 5%. The value $k = 0$ is associated with the null hypothesis, whereas the values $k = 1, \dots, 20$ provide increasingly severe alternatives, the most severe of which involves a first eigenvector that is orthogonal to θ_1^0 . The resulting rejection frequencies are plotted in the right panel of Figure 4. Clearly, they are in perfect agreement with the asymptotic powers, which supports our theoretical results.

We conclude this section by stressing that, as announced in the [Introduction](#), the contiguity rate in Theorem 4.1 depends on the regime considered. Clearly, the weaker the identifiability (that is, the closer the underlying distribution to the spherical Gaussian one), the slower the contiguity rate, that is, the hardest the inference problem on θ_1 .

5. Extension to the elliptical case. Since we focused so far on multinormal distributions, a natural question is whether or not our results extend away from the Gaussian case. In this section, we discuss this in the framework of the most classical extension of multinormal distributions, namely in the class of elliptical distributions. More specifically, we will consider triangular arrays of p -variate observations \mathbf{X}_{ni} , $i = 1, \dots, n$, $n = 1, 2, \dots$, where $\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}$ form a random sample from the p -variate elliptical distribution with location μ_n , covariance matrix $\Sigma_n = \sigma_n(\mathbf{I}_p + r_n v \theta_1 \theta_1')$ (as in (1.2)) and radial density f . That is, we assume that \mathbf{X}_{ni} admits the probability density function (with respect to the Lebesgue measure on \mathbb{R}^p)

$$(5.1) \quad \mathbf{x} \mapsto \frac{c_{p,f}}{(\det \Sigma_n)^{1/2}} f\left(\sqrt{(\mathbf{x} - \mu_n)' \Sigma_n^{-1} (\mathbf{x} - \mu_n)}\right),$$

where $c_{p,f} > 0$ is a normalization factor and where the radial density $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is such that the covariance matrix of \mathbf{X}_{ni} exists and is equal to Σ_n ; f is not a genuine density (as it does not integrate to one), but it determines the density of the Mahalanobis distance

$$d_{ni} := \sqrt{(\mathbf{X}_{ni} - \mu_n)' \Sigma_n^{-1} (\mathbf{X}_{ni} - \mu_n)},$$

which is given by $r \mapsto (\mu_{p-1,f})^{-1} r^{p-1} f(r) \mathbb{I}[r \geq 0]$, with $\mu_{\ell,f} := \int_0^\infty r^\ell f(r) dr$. In this section, we will assume that \mathbf{X}_{ni} , or equivalently d_{ni} , has finite fourth-order moments, that is, we will assume that f belongs to the collection \mathcal{F} of radial densities f above that further satisfy $\mu_{p+3,f} < \infty$. This guarantees finiteness of the elliptical kurtosis coefficient

$$(5.2) \quad \kappa_p(f) := \frac{pE[d_{ni}^4]}{(p+2)(E[d_{ni}^2])^2} - 1 \left(= \frac{p\mu_{p-1,f}\mu_{p+3,f}}{(p+2)\mu_{p+1,f}^2} - 1 \right);$$

see, for example, page 54 of [Anderson \(2003\)](#). Classical radial densities in \mathcal{F} include the Gaussian one $\phi(r) = \exp(-r^2/2)$ or the Student t_ν one $f_\nu(r) = (1 + r^2/(\nu - 2))^{-(p+\nu)/2}$, with $\nu > 4$. The sequence of hypotheses associated with the triangular arrays of observations above will be denoted as $\mathbf{P}_{\mu_n, \sigma_n, \theta_1, r_n, v, f}$.

When it comes to testing the null hypothesis $\mathcal{H}_0^{(n)} : \theta_1 = \theta_1^0$, it is well known that, even in the standard regime (i) ($r_n \equiv 1$), the Anderson test statistic Q_A in (2.1) and the HPV test statistic Q_{HPV} in (2.2) are asymptotically χ_{p-1}^2 under the sequence of null hypotheses $\mathbf{P}_{\mu_n, \sigma_n, \theta_1^0, r_n, v, f}$ if and only if $\kappa_p(f)$ takes the same value $\kappa_p(\phi) = 0$ as in the Gaussian case; see, for example, [Hallin, Paindaveine and Verdebout \(2010\)](#). Consequently, there is no guarantee, even in regime (i), that the corresponding tests ϕ_A and ϕ_{HPV} meet the asymptotic nominal level constraint under ellipticity, and it therefore makes little sense, in the elliptical case, to investigate the robustness of these tests to weak identifiability. This explains why we

will rather focus on their robustified versions ϕ_A^\dagger and ϕ_{HPV}^\dagger , that reject the null hypothesis at asymptotic level α whenever

$$(5.3) \quad Q_A^{(n)\dagger} := \frac{Q_A^{(n)}}{1 + \hat{\kappa}_p^{(n)}} > \chi_{p-1, 1-\alpha}^2 \quad \text{and} \quad Q_{HPV}^{(n)\dagger} := \frac{Q_{HPV}^{(n)}}{1 + \hat{\kappa}_p^{(n)}} > \chi_{p-1, 1-\alpha}^2,$$

respectively, where

$$(5.4) \quad \begin{aligned} \hat{\kappa}_p^{(n)} &:= \frac{p\{\frac{1}{n} \sum_{i=1}^n ((\mathbf{X}_{ni} - \bar{\mathbf{X}}_n)' \mathbf{S}_n^{-1} (\mathbf{X}_{ni} - \bar{\mathbf{X}}_n))^2\}}{(p+2)\{\frac{1}{n} \sum_{i=1}^n (\mathbf{X}_{ni} - \bar{\mathbf{X}}_n)' \mathbf{S}_n^{-1} (\mathbf{X}_{ni} - \bar{\mathbf{X}}_n)\}^2} - 1 \\ &= \frac{1}{np(p+2)} \sum_{i=1}^n ((\mathbf{X}_{ni} - \bar{\mathbf{X}}_n)' \mathbf{S}_n^{-1} (\mathbf{X}_{ni} - \bar{\mathbf{X}}_n))^2 - 1 \end{aligned}$$

is the natural estimator of the kurtosis coefficient $\kappa_p(f)$. In the standard regime (i), Tyler (1981, 1983) showed that ϕ_A^\dagger has asymptotic size α under $P_{\mu_n, \sigma_n, \theta_1^0, r_n, v, f}$, whereas Hallin, Paindaveine and Verdebout (2010) proved the same result for ϕ_{HPV}^\dagger and established the asymptotic equivalence of both tests in probability. In the Gaussian case, these tests are, still in regime (i), asymptotically equivalent to their original versions ϕ_A and ϕ_{HPV} , hence inherit the optimality properties of the latter. The tests ϕ_A^\dagger and ϕ_{HPV}^\dagger may therefore be considered *pseudo-Gaussian* versions of their antecedents, since they extend their validity to the class of elliptical distributions with finite fourth-order moments without sacrificing optimality in the Gaussian case.

The above considerations make it natural to investigate the robustness of these pseudo-Gaussian tests to weak identifiability. Since these tests are invariant under translations and scale transformations, we will still assume, without loss of generality, that $\mu_n \equiv \mathbf{0}$ and $\sigma_n \equiv 1$ (see the discussion below Lemma 2.1), and we will write accordingly $P_{\theta_1, r_n, v, f} := P_{\mathbf{0}, 1, \theta_1, r_n, v, f}$. Note that the Gaussian hypotheses $P_{\theta_1, r_n, v} = P_{\theta_1, r_n, v, \phi}$ are those we considered in the previous sections of the paper. Our results will build on the following elliptical extension of Lemma 2.1.

LEMMA 5.1. *Fix a unit p -vector θ_1^0 , $v > 0$, a bounded positive real sequence (r_n) , and $f \in \mathcal{F}$. Then, under $P_{\theta_1, r_n, v, f}$, $\sqrt{n}(\Sigma_n^{-1/2})^{\otimes 2} \text{vec}(\mathbf{S}_n - \Sigma_n)$ is asymptotically normal with mean zero and covariance matrix $(1 + \kappa_p(f))(\mathbf{I}_{p^2} + \mathbf{K}_p) + \kappa_p(f)(\text{vec } \mathbf{I}_p)(\text{vec } \mathbf{I}_p)'$. In particular, (i) if $r_n \equiv 1$, then $\sqrt{n} \text{vec}(\mathbf{S}_n - \Sigma_n)$ is asymptotically normal with mean zero and covariance matrix $(1 + \kappa_p(f))(\mathbf{I}_{p^2} + \mathbf{K}_p)(\Sigma(v))^{\otimes 2} + \kappa_p(f)(\text{vec } \Sigma(v))(\text{vec } \Sigma(v))'$, still with $\Sigma(v) := \mathbf{I}_p + v\theta_1\theta_1'$; (ii) if r_n is $o(1)$, then $\sqrt{n} \text{vec}(\mathbf{S}_n - \Sigma_n)$ is asymptotically normal with mean zero and covariance matrix $(1 + \kappa_p(f))(\mathbf{I}_{p^2} + \mathbf{K}_p) + \kappa_p(f)(\text{vec } \mathbf{I}_p)(\text{vec } \mathbf{I}_p)'$.*

The main result of this section is the following theorem, that in particular extends Theorem 3.1 to the elliptical setup.

THEOREM 5.1. *Fix a unit p -vector θ_1^0 , $v > 0$, a bounded positive real sequence (r_n) , and $f \in \mathcal{F}$. Then, under $P_{\theta_1^0, r_n, v, f}$,*

$$Q_{HPV}^{(n)\dagger} \xrightarrow{\mathcal{D}} \chi_{p-1}^2,$$

so that, in all regimes (i)–(iv), the test ϕ_{HPV}^\dagger , irrespective of the radial density $f \in \mathcal{F}$, has asymptotic size α under the null hypothesis. Moreover, under $P_{\theta_1^0, r_n, v, \phi}$,

$$(5.5) \quad Q_{HPV}^{(n)\dagger} = Q_{HPV}^{(n)} + o_P(1)$$

as $n \rightarrow \infty$.

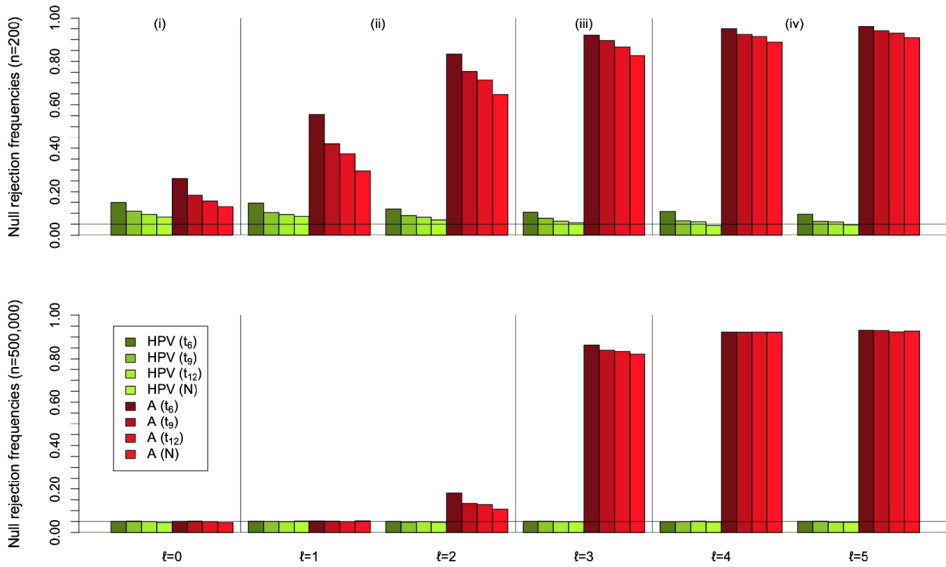


FIG. 5. Empirical rejection frequencies, under the null hypothesis, of the tests $\phi_{\text{HPV}}^\dagger$ and ϕ_A^\dagger performed at nominal level 5%. Results are based on $M = 10,000$ independent ten-dimensional random samples of size $n = 200$ and size $n = 500,000$, drawn from t_6 , t_9 , t_{12} and Gaussian distributions. Increasing values of ℓ bring the underlying spiked covariance matrix closer and closer to a multiple of the identity matrix; see Section 5 for details.

This result shows that the pseudo-Gaussian version $\phi_{\text{HPV}}^\dagger$ of ϕ_{HPV} is robust to weak identifiability under any elliptical distribution with finite fourth-order moments. Since the asymptotic equivalence in (5.5) extends, from contiguity, to the (Gaussian) local alternatives identified in Theorem 4.1, it also directly follows from Theorem 5.1 that $\phi_{\text{HPV}}^\dagger$ inherits the optimality properties of ϕ_{HPV} in the multinormal case. For the sake of completeness, we mention that, by using elliptical extensions of Lemmas 2.2–2.3 (see Lemmas S.4.1–S.4.2 in the Supplementary Material Paindaveine, Remy and Verdebout (2019)), it can be showed that, irrespective of the elliptical distribution considered, the pseudo-Gaussian test ϕ_A^\dagger asymptotically meets the nominal level constraint in regimes (i)–(ii) only, hence is not robust to weak identifiability. Remarkably, by using the same results, it can also be showed that, under any bivariate elliptical distribution with finite fourth-order moments, the null asymptotic distribution of $Q_A^{(n)\dagger}$ is still $4\chi_1^2$ in regime (iv), which extends Corollary 3.1 to the elliptical setup.

We now illustrate these results through a Monte Carlo exercise that extends to the elliptical setup the one conducted in Figure 1. To do so, for any $\ell = 0, 1, \dots, 5$, we generated $M = 10,000$ mutually independent random samples $\mathbf{X}_i^{(\ell,s)}$, $i = 1, \dots, n$, from the ($p = 10$)-variate t_6 ($s = 1$), t_9 ($s = 2$), t_{12} ($s = 3$), and normal ($s = 4$) distributions with mean zero and covariance matrix $\Sigma_n^{(\ell)} := \mathbf{I}_p + n^{-\ell/6}\theta_1^0\theta_1^{0r}$, where θ_1^0 is still the first vector of the canonical basis of \mathbb{R}^p . As in Figure 1, this covers regimes (i) ($\ell = 0$), (ii) ($\ell = 1, 2$), (iii) ($\ell = 3$), and (iv) ($\ell = 4, 5$). Figure 5 reports, for $n = 200$ and $n = 500,000$, the resulting rejection frequencies of the pseudo-Gaussian tests $\phi_{\text{HPV}}^\dagger$ and ϕ_A^\dagger for $\mathcal{H}_0^{(n)} : \theta_1 = \theta_1^0$ at nominal level 5%. Clearly, the results confirm that, irrespective of the underlying elliptical distribution, the pseudo-Gaussian HPV test is robust to weak identifiability, while the pseudo-Gaussian Anderson test meets the asymptotic level constraint only in regimes (i)–(ii) (this test strongly overrejects the null hypothesis in other regimes).

6. Real data example. We now provide a real data illustration on the celebrated Swiss banknote dataset, which has been considered in numerous multivariate statistics monographs, such as Flury and Riedwyl (1988), Atkinson, Riani and Cerioli (2004), Härdle and Simar

(2007) and Koch (2014), but also in many research papers; see, for example, Salibián-Barrera, Van Aelst and Willems (2006) or Burman and Polonik (2009). The dataset, that is available in the R package `uskewfactors` (Murray, Browne and McNicholas (2016)), offers six measurements on 100 genuine and 100 counterfeit old Swiss 1000-franc banknotes. This dataset was often used to illustrate various multivariate statistics procedures such as, for example, linear discriminant analysis (Flury and Riedwyl (1988)), principal component analysis (Flury (1988)), or independent component analysis (Girolami (1999)); we also refer to Shinmura (2016) for a recent account on discriminant analysis for this dataset.

Here, we aim to complement the PCA analysis conducted in Flury (1988) (see pages 41–43), hence use the exact same subset of the Swiss banknote data as the one considered there. More precisely, (i) we focus on four of the six available measurements, namely the width L of the left side of the banknote, the width R on its right side, the width B of the bottom margin and the width T of the top margin, all measured in $\text{mm} \times 10^{-1}$ (rather than in the original mm); (ii) we also restrict to $n = 85$ counterfeit bills made by the same forger (it is well known that the 100 counterfeit bills were made by two different forgers; see, e.g., Flury and Riedwyl (1988), page 250, or Fritz, García-Escudero and Mayo-Iscar (2012), page 22). Letting $c_n = (n - 1)/n \approx 0.99$, the resulting sample covariance matrix is

$$S = c_n \begin{pmatrix} 6.41 & 4.89 & 2.89 & -1.30 \\ 4.89 & 9.40 & -1.09 & 0.71 \\ 2.89 & -1.09 & 72.42 & -43.30 \\ -1.30 & 0.71 & -43.30 & 40.39 \end{pmatrix},$$

with eigenvalues of $\hat{\lambda}_1 = 102.69c_n$, $\hat{\lambda}_2 = 13.05c_n$, $\hat{\lambda}_3 = 10.23c_n$ and $\hat{\lambda}_4 = 2.66c_n$, and corresponding eigenvectors

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

the unimportant factor c_n is used here to ease the comparison with Flury (1988), where the unbiased version of the sample covariance matrix was adopted throughout. From these estimates, Flury concludes that the first principal component is a contrast between B and T , hence can be interpreted as the vertical position of the print image on the bill. It is tempting to interpret the second principal component as an aggregate of L and R , that is, essentially as the vertical size of the bill. Flury, however, explicitly writes “*beware: the second and third roots are quite close and so the computation of standard errors for the coefficients of $\hat{\theta}_2$ and $\hat{\theta}_3$ may be hazardous*”. He reports that these eigenvectors should be considered spherical and that the corresponding standard errors should be ignored. In other words, Flury, due to the structure of the spectrum, refrains from drawing any conclusion about the second component.

The considerations above make it natural to test that L and R contribute equally to the second principal component and that they are the only variables to contribute to it. In other words, it is natural to test the null hypothesis $\mathcal{H}_0 : \theta_2 = \theta_2^0$, with $\theta_2^0 := (1, 1, 0, 0)' / \sqrt{2}$. While the tests discussed in the present paper address testing problems on the first eigenvector θ_1 , obvious modifications of these tests allow performing inference on any eigenvector θ_j , $j =$

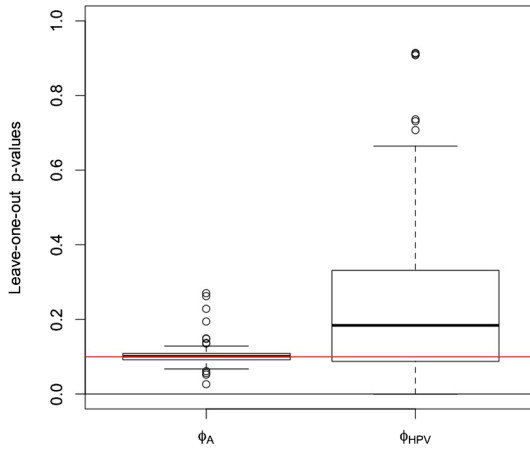


FIG. 6. *Boxplots of the 85 “leave-one-out” p-values of the Anderson test in (6.1) (left) and HPV test in (6.2) (right) when testing the null hypothesis $\mathcal{H}_0 : \theta_2 := (1, 1, 0, 0)' / \sqrt{2}$. More precisely, these p-values are those obtained when applying the corresponding tests to the 85 subsample of size 84 obtained by removing one observation in the real data set considered in the PCA analysis of Flury (1988), pages 41–43.*

$2, \dots, p$. In particular, the Anderson test $\phi_A^{(n)}$ and HPV test $\phi_{HPV}^{(n)}$ for $\mathcal{H}_0 : \theta_2 = \theta_2^0$ against $\mathcal{H}_1 : \theta_2 \neq \theta_2^0$ reject the null hypothesis at asymptotic level α whenever

$$(6.1) \quad n(\hat{\lambda}_2 \theta_2^{0'} \mathbf{S}^{-1} \theta_2^0 + \hat{\lambda}_2^{-1} \theta_2^{0'} \mathbf{S} \theta_2^0 - 2) > \chi_{p-1, 1-\alpha}^2$$

and

$$(6.2) \quad \frac{n}{\hat{\lambda}_2} \sum_{j=1, j \neq 2}^p \hat{\lambda}_j^{-1} (\tilde{\theta}_j' \mathbf{S} \theta_2^0)^2 > \chi_{p-1, 1-\alpha}^2,$$

respectively, where, parallel to (1.1), $\tilde{\theta}_1, \theta_2^0, \tilde{\theta}_3, \dots, \tilde{\theta}_p$ results from a Gram–Schmidt orthogonalization of $\hat{\theta}_1, \theta_2^0, \hat{\theta}_3, \dots, \hat{\theta}_p$. When applied with $\theta_2^0 := (1, 1, 0, 0)' / \sqrt{2}$, this HPV test provides a p -value equal to 0.177, hence does not lead to rejection of the null hypothesis at any usual nominal level. In contrast, the p -value of the Anderson test in (6.1) is 0.099, so that this test rejects the null hypothesis at the level 10%. Since the results of this paper show that the Anderson test tends to strongly overreject the null hypothesis when eigenvalues are close, practitioners should here be confident that the HPV test provides the right decision.

To somewhat assess the robustness of this result, we performed the same HPV and Anderson tests on the 85 subsamples obtained by removing one observation from the sample considered above. For each test, a boxplot of the resulting 85 “leave-one-out” p -values is provided in Figure 6. Clearly, these boxplots reveal that the Anderson test rejects the null hypothesis much more often than the HPV test. Again, the results of the paper provide a strong motivation to rely on the outcome of the HPV test in the present context.

7. Wrap up and perspectives. In this paper, we tackled the problem of testing the null hypothesis $\mathcal{H}_0^{(n)} : \theta_1 = \theta_1^0$ against the alternative $\mathcal{H}_1^{(n)} : \theta_1 \neq \theta_1^0$, where θ_1 is the eigenvector associated with the largest eigenvalue of the underlying covariance matrix and where θ_1^0 is some fixed unit vector. We analyzed the asymptotic behavior of the classical Anderson (1963) test ϕ_A and of the Hallin, Paindaveine and Verdebout (2010) test ϕ_{HPV} under sequences of p -variate Gaussian models with spiked covariance matrices of the form $\Sigma_n = \sigma_n^2(\mathbf{I}_p + r_n v \theta_1 \theta_1')$, where (σ_n) is a positive sequence, $v > 0$ is fixed, and (r_n) is a positive sequence that converges to zero. We showed that in these situations where θ_1 is closer and closer to

being unidentified, ϕ_{HPV} performs better than ϕ_A : (i) ϕ_{HPV} , unlike ϕ_A , meets asymptotically the nominal level constraint without any condition on the rate at which r_n converges to zero, and (ii) ϕ_{HPV} remains locally asymptotically maximin in all regimes, but in the contiguity regime $r_n = 1/\sqrt{n}$ where ϕ_{HPV} still enjoys the same optimality property locally in τ . These considerations, along with the asymptotic equivalence of ϕ_{HPV} and ϕ_A in the standard case $r_n \equiv 1$, clearly imply that the test ϕ_{HPV} , for all practical purposes, should be favored over ϕ_A , all the more so that the results above extend to elliptical distributions if the Anderson and HPV tests are replaced with their pseudo-Gaussian versions ϕ_A^\dagger and $\phi_{\text{HPV}}^\dagger$.

To conclude, we discuss some research perspectives. Throughout the paper, we assumed that the dimension p is fixed. It would be of interest to consider tests that can cope with high-dimensional situations where p is as large as n or even larger than n , and to investigate the robustness of these tests to weak identifiability. The tests considered in the present paper, however, are not suitable in high dimensions. This is clear for the Anderson test ϕ_A since this test requires inverting the sample covariance matrix \mathbf{S}_n , that fails to be invertible for $p \geq n$. As for the HPV test ϕ_{HPV} , our investigation of the asymptotic behavior of this test in the fixed- p case crucially relied on the consistency of the eigenvalues $\hat{\lambda}_{n,j}$ of \mathbf{S} ; in high-dimensional regimes where $p = p_n \rightarrow \infty$ so that $p_n/n \rightarrow c$, however, these sample eigenvalues are no longer consistent (see, e.g., Baik, Ben Arous and P  ch   (2005)), which suggests that ϕ_{HPV} is not robust to high dimensionality. To explore this, we conducted the following Monte Carlo exercise: for $n = 200$ and each value of $p = cn$, with $c \in \{0.5, 0.75, 1, 1.5, 2\}$, we generated 2000 mutually independent random samples $\mathbf{X}_1, \dots, \mathbf{X}_n$ from the p -variate normal distribution with mean zero and covariance matrix $\Sigma = \mathbf{I}_p + \theta_1^0 \theta_1^{0T}$, where θ_1^0 is the first vector of the canonical basis of \mathbb{R}^p . The resulting rejection frequencies of the test ϕ_{HPV} (resp., of the test ϕ_A), conducted at asymptotic level 5%, are 0.9255 (resp., 1) for $c = 0.5$, 0.9240 (resp., 1) for $c = 0.75$, 0.5000 (resp., —) for $c = 1$, 0.1985 (resp., —) for $c = 1.5$, and 0.1715 (resp., —) for $c = 2$ (as indicated above, the Anderson test cannot be used for $p \geq n$). This confirms that neither ϕ_{HPV} nor ϕ_A can cope with high dimensionality. As a result, the problem of providing a suitable test in the high-dimensional setup and of studying its robustness to weak identifiability is widely open and should be investigated in future research.

Acknowledgments. We would like to thank the Editor Edward I. George, the Associate Editor and two anonymous referees for their insightful comments and suggestions that led to a substantial improvement of a previous version of this work.

The first author was supported by a research fellowship from the Francqui Foundation and by the Program of Concerted Research Actions (ARC) of the Universit   libre de Bruxelles.

The third author was supported by the Cr  dit de Recherche J.0134.18 of the FNRS (Fonds National pour la Recherche Scientifique), Communaut   Fran  aise de Belgique and by the aforementioned ARC program of the Universit   libre de Bruxelles.

SUPPLEMENTARY MATERIAL

Supplement to “Testing for principal component directions under weak identifiability” (DOI: [10.1214/18-AOS1805SUPP](https://doi.org/10.1214/18-AOS1805SUPP); .pdf). In this supplement, we prove all theoretical results of the present paper.

REFERENCES

- ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 <https://doi.org/10.1214/aoms/1177704248>
- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR1990662

- ATKINSON, A. C., RIANI, M. and CERIOLI, A. (2004). *Exploring Multivariate Data with the Forward Search*. Springer Series in Statistics. Springer, New York. MR2055967 <https://doi.org/10.1007/978-0-387-21840-3>
- BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 <https://doi.org/10.1214/009117905000000233>
- BALI, J. L., BOENTE, G., TYLER, D. E. and WANG, J.-L. (2011). Robust functional principal components: A projection-pursuit approach. *Ann. Statist.* **39** 2852–2882. MR3012394 <https://doi.org/10.1214/11-AOS923>
- BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 <https://doi.org/10.1214/13-AOS1127>
- BOENTE, G. and FRAIMAN, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.* **48** 335–345. MR1771495 [https://doi.org/10.1016/S0167-7152\(00\)00014-6](https://doi.org/10.1016/S0167-7152(00)00014-6)
- BURMAN, P. and POLONIK, W. (2009). Multivariate mode hunting: Data analytic tools with measures of significance. *J. Multivariate Anal.* **100** 1198–1218. MR2508381 <https://doi.org/10.1016/j.jmva.2008.10.015>
- CROUX, C. and HAESBROECK, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika* **87** 603–618. MR1789812 <https://doi.org/10.1093/biomet/87.3.603>
- CUEVAS, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference* **147** 1–23. MR3151843 <https://doi.org/10.1016/j.jspi.2013.04.002>
- DUFOUR, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* **65** 1365–1387. MR1604305 <https://doi.org/10.2307/2171740>
- DUFOUR, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *J. Econometrics* **133** 443–477. MR2252905 <https://doi.org/10.1016/j.jeconom.2005.06.007>
- FLURY, B. (1988). *Common Principal Components and Related Multivariate Models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, New York. MR0986245 https://doi.org/10.1007/978-3-642-01932-6_28
- FLURY, B. and RIEDWYL, H. (1988). *Multivariate Statistics: A Practical Approach*. CRC Press, London.
- FORCHINI, G. and HILLIER, G. (2003). Conditional inference for possibly unidentified structural equations. *Econometric Theory* **19** 707–743. MR2002577 <https://doi.org/10.1017/S0266466603195011>
- FRITZ, H., GARCÍA-ESCUADERO, L. A. and MAYO-ISCAR, A. (2012). tclust: An R package for a trimming approach to cluster analysis. *J. Stat. Softw.* **47**.
- GIROLAMI, M. (1999). *Self-Organizing Neural Networks. Independent Component Analysis and Blind Source Separation*. Springer, London.
- HALLIN, M. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465 <https://doi.org/10.1214/009053606000000731>
- HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38** 3245–3299. MR2766852 <https://doi.org/10.1214/10-AOS810>
- HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2014). Efficient R-estimation of principal and common principal components. *J. Amer. Statist. Assoc.* **109** 1071–1083. MR3265681 <https://doi.org/10.1080/01621459.2014.880057>
- HAN, F. and LIU, H. (2014). Scale-invariant sparse PCA on high-dimensional meta-elliptical data. *J. Amer. Statist. Assoc.* **109** 275–287. MR3180563 <https://doi.org/10.1080/01621459.2013.844699>
- HÄRDLE, W. and SIMAR, L. (2007). *Applied Multivariate Statistical Analysis*, 2nd ed. Springer, Berlin. MR2367300
- HE, R., HU, B.-G., ZHENG, W.-S. and KONG, X.-W. (2011). Robust principal component analysis based on maximum correntropy criterion. *IEEE Trans. Image Process.* **20** 1485–1494. MR2828599 <https://doi.org/10.1109/TIP.2010.2103949>
- HUBERT, M., ROUSSEEUW, P. J. and VANDEN BRANDEN, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics* **47** 64–79. MR2135793 <https://doi.org/10.1198/004017004000000563>
- JACKSON, J. E. (2005). *A User's Guide to Principal Components*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ.
- JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11** 818–887. MR1458943 <https://doi.org/10.1017/S0266466600009907>
- JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 <https://doi.org/10.1198/jasa.2009.0121>
- JOLICOEUR, P. (1984). Principal components, factor analysis, and multivariate allometry: A small-sample direction test. *Biometrics* **40** 685–690.

- KOCH, I. (2014). *Analysis of Multivariate and High-Dimensional Data*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press, New York. MR3154467
- MAGNUS, J. R. and NEUDECKER, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Wiley, Chichester. MR1698873
- MURRAY, P. M., BROWNE, R. P. and MCNICHOLAS, P. D. (2016). uskewFactors: Model-based clustering via mixtures of unrestricted skew-t sactor analyzer models. R package. Available at <https://cran.r-project.org/web/packages/uskewFactors/index.html>.
- PAINDAVEINE, D., REMY, J. and VERDEBOUT, T. (2019). Supplement to “Testing for principal component directions under weak identifiability.” <https://doi.org/10.1214/18-AOS1805SUPP>.
- PAINDAVEINE, D. and VERDEBOUT, T. (2017). Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities. *Ann. Statist.* **45** 800–832. MR3650401 <https://doi.org/10.1214/16-AOS1468>
- PÖTSCHER, B. M. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* **70** 1035–1065. MR1910411 <https://doi.org/10.1111/1468-0262.00318>
- ROUSSAS, G. G. and BHATTACHARYA, D. (2011). Revisiting local asymptotic normality (LAN) and passing on to local asymptotic mixed normality (LAMN) and local asymptotic quadratic (LAQ) experiments. In *Advances in Directional and Linear Statistics* (M. T. Wells and A. Sengupta, eds.) 253–280. Physica-Verlag/Springer, Heidelberg. MR2767545 https://doi.org/10.1007/978-3-7908-2628-9_17
- SALIBIÁN-BARRERA, M., VAN AELST, S. and WILLEMS, G. (2006). Principal components analysis based on multivariate MM estimators with fast and robust bootstrap. *J. Amer. Statist. Assoc.* **101** 1198–1211. MR2328307 <https://doi.org/10.1198/016214506000000096>
- SCHWARTZMAN, A., MASCARENHAS, W. F. and TAYLOR, J. E. (2008). Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann. Statist.* **36** 2886–2919. MR2485016 <https://doi.org/10.1214/08-AOS628>
- SHINMURA, S. (2016). *New Theory of Discriminant Analysis After R. Fisher*. Springer, Singapore. MR3617625 <https://doi.org/10.1007/978-981-10-2164-0>
- SYLVESTER, A. D., KRAMER, P. A. and JUNGERS, W. L. (2008). Modern humans are not (quite) isometric. *Amer. J. Phys. Anthropol.* **137** 371–383.
- TYLER, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736. MR0619278
- TYLER, D. E. (1983). A class of asymptotic tests for principal component vectors. *Ann. Statist.* **11** 1243–1250. MR0720269 <https://doi.org/10.1214/aos/1176346337>
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics **3**. Cambridge Univ. Press, Cambridge. MR1652247 <https://doi.org/10.1017/CBO9780511802256>