High-Dimensional Tests for Spherical Location and Spiked Covariance

Christophe Ley*, Davy Paindaveine*,[1] and Thomas Verdebout†

* *Département de Mathématique and ECARES*
Université Libre de Bruxelles
Avenue F.D. Roosevelt, 50 - ECARES, CP 114/04
and Campus Plaine, CP 210, bvd du triomphe
B-1050 Bruxelles, BELGIUM

† *EQUIPPE and INRIA*
Université Lille III
Domaine Universitaire du Pont de Bois, BP 60149
F-59653 Villeneuve d'Ascq Cedex, FRANCE

Abstract

This paper mainly focusses on one of the most classical testing problems in directional statistics, namely the spherical location problem that consists in testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ under which the (rotational) symmetry center $\boldsymbol{\theta}$ is equal to a given value $\boldsymbol{\theta}_0$. The most classical procedure for this problem is the so-called Watson test, which is based on the sample mean of the observations. This test enjoys many desirable properties, but its asymptotic theory requires the sample size n to be large compared to the dimension p. This is a severe limitation, since more and more problems nowadays involve high-dimensional directional data (e.g., in genetics or text mining). In the present work, we derive the asymptotic null distribution of the Watson statistic as both n and p go to infinity. This reveals that (i) the Watson test is robust against high dimensionality, and that (ii) it allows for (n, p)-asymptotic results that are universal, in the sense that p may go to infinity arbitrarily fast (or slowly) as a function of n. Turning to Euclidean data, we show that our results also lead to a test for the null that the covariance matrix of a high-dimensional multinormal distribution has a "$\boldsymbol{\theta}_0$-spiked" structure. Finally, Monte Carlo studies corroborate our asymptotic results and briefly explore non-null rejection frequencies.

Keywords: Directional statistics, high-dimensional data, location tests, principal component analysis, rotationally symmetric distributions, spherical mean

1. Introduction

The technological advances and the ensuing new devices to collect and store data lead nowadays in many disciplines to data sets with very high dimension p, often larger than the sample size n. Consequently, there is a need for inferential methods that can deal with such high-dimensional data, and this has entailed a huge activity related to high-dimensional problems in the last decade. One- and multi-sample location problems have been investigated in Srivastava and Fujikoshi (2006), Schott (2007), Chen and Qin (2010),

[1]Corresponding author: dpaindav@ulb.ac.be

Srivastava et al. (2013), and Srivastava and Kubokawa (2013), among others. Since the seminal paper Ledoit and Wolf (2002), problems related to covariance or scatter matrices have also been thoroughly studied by several authors; see, e.g., Chen et al. (2010), Li and Chen (2012), Onatski et al. (2013) and Jiang and Yang (2013). In particular, the problem of testing for sphericity has attracted much attention.

In this paper, we are interested in high-dimensional *directional* data, that is, in data lying on the unit hypersphere

$$\mathcal{S}^{p-1} = \Big\{ \mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}} = 1 \Big\},$$

with p large. Such data occur when only the direction of the observations and not their magnitude matters, and are extremely common, e.g., in magnetic resonance (Dryden, 2005), gene-expression (Banerjee et al., 2003), and text mining (Banerjee et al., 2005). Inference for high-dimensional directional data has already been considered in several papers. For instance, Banerjee and Ghosh (2002, 2004) and Banerjee et al. (2005) investigate clustering methods in this context. Most asymptotic results available, however, have been obtained as p goes to infinity, with n fixed. This is the case of almost all results in Stam (1982), Watson (1983a), Watson (1988), and Dryden (2005). To the best of our knowledge, the only (n, p)-asymptotic results available can be found in Dryden (2005), Cai and Jiang (2012), Cai et al. (2013), and Paindaveine and Verdebout (2015a). However, Dryden (2005) imposes the stringent condition that $p/n^2 \to \infty$ when studying the asymptotic behavior of the classical pseudo-FvML location estimator (FvML here refers to *Fisher-von Mises-Langevin* distributions; see below). Cai and Jiang (2012) and Cai et al. (2013) consider various (n, p)-asymptotic regimes in the context of testing for uniformity on the unit sphere, but the tests to be used depend on the regime considered which makes practical implementation problematic. Finally, Paindaveine and Verdebout (2015a) propose tests that are robust to the (n, p)-asymptotic regime considered; their tests, however, are sign procedures, hence are not based on sufficient statistics — unlike the much more classical pseudo-FvML procedures.

In the present paper, we intend to overcome these limitations in the context of the spherical location problem, one of the most fundamental problems in directional statistics. The natural distributional framework for this problem is provided by the class of *rotationally symmetric distributions* (see Section 2), that is a semiparametric model, indexed by a finite-dimensional (location) parameter $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$ and an infinite-dimensional parameter F. The spherical location problem is the problem

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ \mathcal{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \end{cases}$$

where $\boldsymbol{\theta}_0$ is a given unit vector and F remains unspecified. The classical test for this problem is the so-called Watson test, based on the sample mean of the observations; see Watson (1983b). This test enjoys many desirable properties, and in particular is a *pseudo-FvML* procedure : in other words, it achieves optimality under FvML distributions, yet remains valid (in the sense that it meets the asymptotic nominal level constraint) under extremely mild assumptions on F.

Unfortunately, nothing is known about the validity of the Watson test in the high-dimensional setup, which, in view of the growing number of high-dimensional directional data to be analyzed, is a severe limitation. Therefore, the aim of this paper is to investigate this issue. We derive the (n, p)-asymptotic null properties of the Watson test. Our results require minimal distributional assumptions and allow for virtually any rotationally symmetric distributions. Even better : in contrast with earlier asymptotic investigations of high-dimensional pseudo-FvML procedures, our asymptotic results are "universal" in the sense that they only require that p goes to infinity as n does (p may go arbitrarily fast (or slowly) to infinity as a function of n). Moreover, as an interesting by-product, we show that our procedures can be used to test the null hypothesis that the covariance matrix of a high-dimensional multinormal distribution is "$\boldsymbol{\theta}_0$-spiked", meaning that it is of the form $\boldsymbol{\Sigma} = \sigma^2(\mathbf{I}_p + \lambda \boldsymbol{\theta}_0 \boldsymbol{\theta}_0')$ for some $\sigma^2, \lambda > 0$ and $\boldsymbol{\theta}_0 \in \mathbb{R}^k$; see, e.g., Johnstone (2001) or the quite recent Onatski et al. (2013) where this covariance structure has been used as an alternative to sphericity.

The outline of the paper is as follows. In Section 2, we define the class of rotationally symmetric distributions and introduce the Watson test for spherical location. In Section 3, we propose a standardized Watson test statistic and derive its asymptotic null distribution in the high-dimensional setting. We also prove that, in some cases, it is asymptotically equivalent to a sign test statistic. In Section 4, we show that the standardized Watson test further allows to test for a spiked covariance structure in high-dimensional multinormal distributions. Monte Carlo studies are conducted in Section 5, while an Appendix collects the proofs of some technical lemmas.

2. Rotational symmetry and the Watson test

The distribution of the random p-vector $\mathbf{X}$, with values on the unit hypersphere $\mathcal{S}^{p-1}$, is *rotationally symmetric* about location $\boldsymbol{\theta}(\in \mathcal{S}^{p-1})$ if $\mathbf{OX}$ is equal in distribution to $\mathbf{X}$ for any orthogonal $p \times p$ matrix $\mathbf{O}$ satisfying $\mathbf{O}\boldsymbol{\theta} = \boldsymbol{\theta}$; see Saw (1978). Rotationally symmetric distributions are characterized by the location parameter $\boldsymbol{\theta}$ and an infinite-dimensional parameter, the cumulative distribution function F of $\mathbf{X}'\boldsymbol{\theta}$, hence they are of a semiparametric nature. The rotationally symmetric distribution associated with $\boldsymbol{\theta}$ and F will be denoted as $\mathcal{R}(\boldsymbol{\theta}, F)$ in the sequel. The most celebrated members of this family are the Fisher-von Mises-Langevin distributions, corresponding to

$$F_{p,\kappa}(t) = c_{p,\kappa} \int_{-1}^{t} (1 - s^2)^{(p-3)/2} \exp(\kappa s)\, ds \qquad (t \in [-1, 1]),$$

where $c_{p,\kappa}$ is a normalization constant and $\kappa(> 0)$ is a *concentration* parameter (the larger the value of κ, the more concentrated about $\boldsymbol{\theta}$ the distribution is); see Mardia and Jupp (2000) for further details.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from $\mathcal{R}(\boldsymbol{\theta}, F)$ and consider the problem of testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $\mathcal{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0 \in \mathcal{S}^{p-1}$ is fixed and F remains unspecified. At first sight, the rotational symmetry assumption about $\boldsymbol{\theta}_0$ may appear quite restrictive. Note however that it contains the null hypothesis of uniformity on the sphere, which itself contains the null hypothesis of sphericity for Euclidean data (since the uniform distribution on the sphere may be obtained by projecting spherical distributions on the sphere), a null that has been the topic of numerous papers in high-dimensional statistics.

Letting $\bar{\mathbf{X}} := \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$, the classical test for the problem above rejects the null for large values of the Watson statistic

$$W_n := \frac{n(p-1)\bar{\mathbf{X}}'(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}_0')\bar{\mathbf{X}}}{1-\frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i'\boldsymbol{\theta}_0)^2}\,. \tag{2.1}$$

Under very mild assumptions on F, the fixed-p asymptotic null distribution of W_n is chi-square with $p-1$ degrees of freedom. The resulting test, ϕ_n^W say, therefore rejects the null, at asymptotic level α, whenever $W_n > \Psi_{p-1}^{-1}(1-\alpha)$, where Ψ_{p-1} stands for the cumulative distribution function of the chi-square distribution with $p-1$ degrees of freedom; see Watson (1983b).

Beyond achieving asymptotic level α under virtually any rotationally symmetric distribution, ϕ_n^W is optimal — more precisely, locally and asymptotically maximin, in the Le Cam sense — when the underlying distribution is FvML; for details, we refer to Paindaveine and Verdebout (2015b), where the asymptotic properties of ϕ_n^W under local alternatives are derived. Although ϕ_n^W is based on the sample mean of the observations, these excellent power properties are not obtained at the expense of robustness, since observations by construction are on the unit hypersphere.

Consequently, ϕ_n^W is a nice solution to the testing problem considered on all counts but one : implementation is based on fixed-p asymptotics, so that ϕ_n^W cannot be used when p is of the same order as, or even larger than, n. The goal of the present work is therefore to investigate the (n,p)-asymptotic properties of the Watson test. We will show that, as n and p go to infinity, the standardized Watson test statistic

$$\tilde{W}_n := \frac{W_n-(p_n-1)}{\sqrt{2(p_n-1)}} \tag{2.2}$$

is asymptotically normal under the null. This of course leads to a high-dimensional Watson test that consists in rejecting the null, at asymptotic level α, whenever $\tilde{W}_n$ exceeds the upper α-quantile of the standard normal distribution. This test clearly is asymptotically equivalent to the original (fixed-p) Watson test based on chi-square critical values, so that the latter may be considered robust to high dimensionality.

3. A high-dimensional Watson test

Consider the high-dimensional version of the testing problem $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $\mathcal{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, based on a triangular array of observations $\mathbf{X}_{ni}$, $i = 1,\ldots,n$, $n = 1,2,\ldots$, where $\mathbf{X}_{ni}$ takes values in $\mathcal{S}^{p_n-1}$ and p_n goes to infinity with n. Using the (null) *tangent-normal decomposition* $\mathbf{X}_{ni} = (\mathbf{X}_{ni}'\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 + v_{ni}\mathbf{S}_{ni}$, where

$$v_{ni} := \|\mathbf{X}_{ni} - (\mathbf{X}_{ni}'\boldsymbol{\theta}_0)\boldsymbol{\theta}_0\| = \sqrt{1-(\mathbf{X}_{ni}'\boldsymbol{\theta}_0)^2}$$

and

$$\mathbf{S}_{ni} := \begin{cases} \dfrac{\mathbf{X}_{ni} - (\mathbf{X}'_{ni}\boldsymbol{\theta}_0)\boldsymbol{\theta}_0}{\|\mathbf{X}_{ni} - (\mathbf{X}'_{ni}\boldsymbol{\theta}_0)\boldsymbol{\theta}_0\|} & \text{if } \mathbf{X}_{ni} \neq \boldsymbol{\theta}_0 \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

the Watson statistic rewrites

$$\begin{aligned} W_n = \frac{p_n - 1}{\sum_{i=1}^n v_{ni}^2} \sum_{i,j=1}^n v_{ni} v_{nj} \mathbf{S}'_{ni} \mathbf{S}_{nj} &= \frac{p_n - 1}{\sum_{i=1}^n v_{ni}^2} \left(\sum_{i=1}^n v_{ni}^2 + 2 \sum_{1 \leq i < j \leq n} v_{ni} v_{nj} \mathbf{S}'_{ni} \mathbf{S}_{nj} \right) \\ &= (p_n - 1) + \frac{2(p_n - 1)}{\sum_{i=1}^n v_{ni}^2} \sum_{1 \leq i < j \leq n} v_{ni} v_{nj} \mathbf{S}'_{ni} \mathbf{S}_{nj}. \end{aligned}$$

The standardized Watson statistic in (2.2) then takes the form

$$\tilde{W}_n = \frac{\sqrt{2(p_n - 1)}}{\sum_{i=1}^n v_{ni}^2} \sum_{1 \leq i < j \leq n} v_{ni} v_{nj} \mathbf{S}'_{ni} \mathbf{S}_{nj}. \tag{3.3}$$

The following result provides the (n,p)-asymptotic null distribution of $\tilde{W}_n$ (see the Appendix for the proof).

Theorem 3.1. *Let $\mathbf{X}_{ni}$, $i = 1, \ldots, n$, $n = 1, 2, \ldots$, form a triangular array of random vectors satisfying the following conditions : (i) for any n, $\mathbf{X}_{n1}, \mathbf{X}_{n2}, \ldots, \mathbf{X}_{nn}$ are mutually independent and share a common rotationally symmetric distribution on $\mathcal{S}^{p_n - 1}$ with location $\boldsymbol{\theta}_0$; (ii) $p_n \to \infty$ as $n \to \infty$; (iii) $\mathrm{E}[v_{n1}^2] > 0$ for any n; (iv) $\mathrm{E}[v_{n1}^4]/(\mathrm{E}[v_{n1}^2])^2 = o(n)$ as $n \to \infty$. Then $\tilde{W}_n$ is asymptotically standard normal.*

The assumptions of Theorem 3.1 are extremely mild. Note in particular that it is not assumed that the common distribution of the $\mathbf{X}_{ni}$'s is absolutely continuous with respect to the surface area measure on $\mathcal{S}^{p_n-1}$. Assumption (iii) only excludes the degenerate case for which $\mathbf{X}_{n1} = \boldsymbol{\theta}_0$ almost surely, which would imply that W_n — hence also $\tilde{W}_n$ — is not well-defined. Most importantly, it should be noted that Assumption (ii) allows p_n to go to infinity in an arbitrary way with n, so that Theorem 3.1 provides a "(n,p)-universal" asymptotic distribution result for the standardized Watson statistic.

Assumption (iv) possibly looks more stringent. However, a sufficient (yet not necessary) condition for (iv) is that $\sqrt{n}\,\mathrm{E}[v_{n1}^2] \to \infty$ as $n \to \infty$. In other words, if (iv) does not hold, we must then have that, for some constant $C > 0$,

$$\mathrm{E}[(\mathbf{X}'_{n1}\boldsymbol{\theta}_0)^2] \geq 1 - \frac{C}{\sqrt{n}} \tag{3.4}$$

for infinitely many n. In the high-dimensional setup considered, (3.4) is extremely pathological, since it corresponds to the distribution of $\mathbf{X}_{n1}$ concentrating in *one* particular direction — namely, the direction $\boldsymbol{\theta}_0$ — in the expanding Euclidean space $\mathbb{R}^{p_n}$. Moreover, there are parametric classes of distributions on the sphere for which Assumption (iv) always holds. An important example is the class of FvML distributions. To show this, note that the integral representation

$$\mathcal{I}_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma(\nu+\frac{1}{2})} \int_{-1}^{1} (1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\, ds$$

of the modified Bessel function of the first kind $\mathcal{I}_\nu(z)$ (see, e.g., Watson (1944), Page 79) readily yields

$$c_{p,\kappa}(\ell) := \int_{-1}^{1} (1-s^2)^{(p+\ell-3)/2} \exp(\kappa s)\, ds = \frac{\sqrt{\pi}\,\Gamma(\frac{p+\ell-1}{2})\mathcal{I}_{\frac{p+\ell}{2}-1}(\kappa)}{(\kappa/2)^{\frac{p+\ell}{2}-1}} \cdot$$

for any nonnegative integer ℓ. If $\mathbf{X}_{1n}$ follows an FvML distribution with a concentration κ_n that is allowed to depend on the sample size n, then

$$\mathrm{E}[v_{n1}^\ell] = \mathrm{E}[(1-(\mathbf{X}_{n1}'\boldsymbol{\theta}_0)^2)^{\ell/2}] = \frac{c_{p_n,\kappa_n}(\ell)}{c_{p_n,\kappa_n}(0)} = \frac{\Gamma(\frac{p_n+\ell-1}{2})\mathcal{I}_{\frac{p_n+\ell}{2}-1}(\kappa_n)}{(\kappa_n/2)^{\frac{\ell}{2}}\Gamma(\frac{p_n-1}{2})\mathcal{I}_{\frac{p_n}{2}-1}(\kappa_n)},$$

which, by using the log-concavity (for any fixed κ) of $\nu \mapsto \mathcal{I}_\nu(\kappa)$ (see, e.g., Baricz and Ponnusamy (2013)) and the identity $\Gamma(z+1) = z\Gamma(z)$, yields

$$\frac{\mathrm{E}[v_{n1}^4]}{(\mathrm{E}[v_{n1}^2])^2} = \frac{(p_n+1)\mathcal{I}_{\frac{p_n}{2}+1}(\kappa_n)\mathcal{I}_{\frac{p_n}{2}-1}(\kappa_n)}{(p_n-1)(\mathcal{I}_{\frac{p_n}{2}}(\kappa_n))^2} \leq \frac{p_n+1}{p_n-1} \leq 3.$$

Consequently, Assumption (iv) is fulfilled in the FvML case, irrespective of the dependence of (κ_n, p_n) in n—hence, also if κ_n goes to infinity arbitrarily fast. On all counts, thus, Assumption (iv) is extremely mild, too.

Theorem 3.1 states that the standardized Watson test statistic $\tilde{W}_n$ is asymptotically standard normal under the null. It is natural to try and control how much the cumulative distribution function of $\tilde{W}_n$ deviates from normality. This can be achieved by using the main result from Heyde and Brown (1970) and leads to the following theorem (see the Appendix for the proof).

Theorem 3.2. *Let $\mathbf{X}_{ni}$, $i = 1, \ldots, n$, $n = 1, 2, \ldots$, form a triangular array of random vectors satisfying the following conditions : (i) for any n, $\mathbf{X}_{n1}, \mathbf{X}_{n2}, \ldots, \mathbf{X}_{nn}$ are mutually independent and share a common rotationally symmetric distribution on $\mathcal{S}^{p_n-1}$ with location $\boldsymbol{\theta}_0$; (ii) $\mathrm{E}[v_{n1}^2] > 0$ for any n; (iii) $\mathrm{E}[v_{n1}^4]/(\mathrm{E}[v_{n1}^2])^2 = o(n)$ as $n \to \infty$. Let*

$$\tilde{\tilde{W}}_n = \left(\frac{n}{n-1}\right)^{1/2} \tilde{W}_n.$$

Then there exists a positive constant C such that, for n large enough,

$$\sup_{z\in\mathbb{R}} \big|\mathrm{P}\big[\tilde{\tilde{W}}_n \leq z\big] - \Phi(z)\big| \leq C\left(\frac{\mathrm{E}[v_{n1}^4]}{n(\mathrm{E}[v_{n1}^2])^2} + \frac{1}{p_n}\right)^{1/5},$$

where Φ denotes the cumulative distribution function of the standard normal distribution.

Of course, if it is further assumed that $p_n \to \infty$ as $n \to \infty$, then this yields Theorem 3.1

(uniformity is no reinforcement here since the limiting distribution is continuous). More importantly, if more stringent assumptions are imposed on $\mathrm{E}[v_{n1}^4]/(\mathrm{E}[v_{n1}^2])^2$ and p_n, then Theorem 3.2 further provides (uniform) rates of convergence. For instance, if it is assumed that $\mathrm{E}[v_{n1}^4]/(\mathrm{E}[v_{n1}^2])^2 = O(1)$ and $1/p_n = O(1/n)$, then Theorem 3.2 yields that $\sup_{z\in\mathbb{R}} \big|\mathrm{P}[\tilde{\tilde{W}}_n \le z] - \Phi(z)\big| = O(n^{-1/5})$ as $n \to \infty$. Clearly, non-trivial convergence rates can only be obtained by imposing a minimal rate at which p_n should go to infinity, which is incompatible with the "universal asymptotics phenomenon" we describe in this paper. We therefore do not pursue this direction in the sequel.

Theorems 3.1-3.2 lead to the test announced at the end of Section 2, namely the test, $\tilde{\phi}_n^W$ say, that rejects the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ in favor of $\mathcal{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ at asymptotic level α whenever

$$\tilde{W}_n > \Phi^{-1}(1-\alpha).$$

As usual, these tests can be inverted to obtain a confidence zone for the symmetry center $\boldsymbol{\theta}$. More precisely, denoting by $\tilde{W}_n(\boldsymbol{\theta}_0)$ the high-dimensional Watson test statistic for the null $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the region

$$R_n = \Big\{\boldsymbol{\theta} \in \mathcal{S}^{p_n-1} : \tilde{W}_n(\boldsymbol{\theta}) \le \Phi^{-1}(1-\alpha)\Big\}$$

is an (n,p)-asymptotically valid confidence zone for $\boldsymbol{\theta}$. Of course, from a practical point of view, one needs to be able to determine R_n, which may be computationally challenging. This problem, that was not even considered for small p, is beyond the scope of this paper.

We stress that both the high-dimensional tests and confidence zones above are asymptotically valid in a "universal" way, that is, irrespective of the way p_n goes to infinity with n. In particular, this implies that the original (fixed-p) test ϕ_n^W, that is asymptotically equivalent to $\tilde{\phi}_n^W$, is asymptotically valid in the high-dimensional case, hence is robust to high dimensionality.

Finally, for the testing problem considered above, Paindaveine and Verdebout (2015a) introduced the high-dimensional *sign* statistic

$$\tilde{Q}_n := \frac{\sqrt{2(p_n-1)}}{n} \sum_{1\le i<j\le n} \mathbf{S}_{ni}'\mathbf{S}_{nj} \tag{3.5}$$

and showed that the (n,p)-universal asymptotic null distribution of $\tilde{Q}_n$ is standard normal. In the next result (that is also proved in the Appendix), we identify assumptions on the sequence (v_{n1}) under which $\tilde{W}_n$ and $\tilde{Q}_n$ are ((n,p)-universally) asymptotically equivalent in probability under the null.

Theorem 3.3. *Let the assumptions of Theorem 3.1 hold and further assume that (v)* $\mathrm{E}[v_{n1}^2]/(\mathrm{E}[v_{n1}])^2 \to 1$ *as* $n \to \infty$. *Then,* $\tilde{W}_n - \tilde{Q}_n = o_{\mathrm{P}}(1)$ *as* $n \to \infty$.

This result shows that, quite intuitively, if v_{n1} becomes constant asymptotically (in the sense that $\mathrm{Var}[v_{n1}]/(\mathrm{E}[v_{n1}])^2 \to 0$), then the high-dimensional Watson test $\tilde{\phi}_n^W$ coincides with the sign test based on (3.5). Note, however, that there is no particular reason why the distribution of $\mathbf{X}_{n1}$ should concentrate on the intersection of the sphere with (a possibly translated version of) the orthogonal complement of $\boldsymbol{\theta}_0$.

4. Spiked covariance matrices

Let $\mathbf{Y}_{n1}, \ldots, \mathbf{Y}_{nn}$ be a random sample from the p_n-dimensional multinormal distribution with mean zero and covariance matrix $\mathbf{\Sigma}$. For fixed $\boldsymbol{\theta}_0 \in \mathcal{S}^{p_n-1}$, we consider here the problem of testing the null hypothesis that $\mathbf{\Sigma}$ has a "$\boldsymbol{\theta}_0$-spiked" structure, that is, is of the form

$$\mathcal{H}_0^{\mathrm{spi}} : \mathbf{\Sigma} = \sigma^2(\mathbf{I}_{p_n} + \lambda \boldsymbol{\theta}_0 \boldsymbol{\theta}_0'), \quad \text{for some } \sigma^2 > 0 \text{ and } \lambda \geq 0.$$

Consider the projections $\mathbf{X}_{ni} := \mathbf{Y}_{ni}/\|\mathbf{Y}_{ni}\|$, $i = 1, \ldots, n$, of the observations on the unit hypersphere, and let

$$\mathbf{S}_{ni} := \frac{\mathbf{X}_{ni} - (\mathbf{X}_{ni}'\boldsymbol{\theta}_0)\boldsymbol{\theta}_0}{\|\mathbf{X}_{ni} - (\mathbf{X}_{ni}'\boldsymbol{\theta}_0)\boldsymbol{\theta}_0\|}.$$

Under $\mathcal{H}_0^{\mathrm{spi}}$, (i) the $\mathbf{S}_{ni}$'s are mutually independent and are uniformly distributed over $\mathcal{S}^{p_n-1}(\boldsymbol{\theta}_0^{\perp}) := \{\mathbf{x} \in \mathcal{S}^{p_n-1} \,|\, \mathbf{x}'\boldsymbol{\theta}_0 = 0\}$; moreover, (ii) the $\mathbf{X}_{ni}'\boldsymbol{\theta}_0$'s are independent and identically distributed, and they are independent of the $\mathbf{S}_{ni}$'s. It is well-known that (i)-(ii) imply that the common distribution of the projected observations $\mathbf{X}_{ni}$ is rotationally symmetric about $\boldsymbol{\theta}_0$. Consequently, a high-dimensional test for $\boldsymbol{\theta}_0$-spikedness is the test, $\tilde{\phi}_n^{\mathrm{spi}}$ say, that rejects the null $\mathcal{H}_0^{\mathrm{spi}}$, at asymptotic level α, whenever

$$\tilde{W}_n^{\mathrm{spi}}(\mathbf{Y}_{n1}, \ldots, \mathbf{Y}_{nn}) := \tilde{W}_n(\mathbf{X}_{n1}, \ldots, \mathbf{X}_{nn}) > \Phi^{-1}(1-\alpha).$$

Theorem 3.1 ensures that $\tilde{\phi}_n^{\mathrm{spi}}$ has asymptotic null size α as soon as p_n goes to infinity with n (universal (n,p)-asymptotics), which is illustrated in the simulations of the next section. Typically, this test will show large powers against $\boldsymbol{\theta}$-spiked alternatives, with $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ and $\lambda > 0$.

5. Monte Carlo simulation study

5.1. Null behavior

In this section, our aim is to check the validity of our universal asymptotic results related to both $\tilde{W}_n$ and $\tilde{W}_n^{\mathrm{spi}}$. To do so, we generated, for every $(n,p) \in C \times C$, with $C = \{5, 30, 200, 1\,000\}$, and with $\boldsymbol{\theta}_0$ the first vector of the canonical basis of $\mathbb{R}^p$, $M = 2\,500$ independent random samples from each of the following p-dimensional distributions :

(i) the FvML distribution $\mathcal{R}(\boldsymbol{\theta}_0, F_{p,2})$ (see Section 2);

(ii) the Purkayastha distribution $\mathcal{R}(\boldsymbol{\theta}_0, G_{p,1})$, associated with

$$G_{p,\kappa}(t) = d_{p,\kappa} \int_{-1}^{t} (1-s^2)^{(p-3)/2} \exp(-\kappa \arccos(s))\, ds \qquad (t \in [-1,1]),$$

where $d_{p,\kappa}$ is a normalizing constant;

(iii) the multinormal distribution with mean zero and covariance matrix $\mathbf{\Sigma} = \mathbf{I}_p + (1/2)\boldsymbol{\theta}_0\boldsymbol{\theta}_0'$.

The standardized Watson statistic $\tilde{W}_n$ was evaluated on the samples from (i)-(ii) (rotational symmetry about $\boldsymbol{\theta}_0$), while the statistic $\tilde{W}_n^{\mathrm{spi}}$ was computed for each sample from (iii) ($\boldsymbol{\theta}_0$-spikedness). For each (n,p)-regime considered, we report the corresponding histograms of $\tilde{W}_n$ in Figures 1-2 and those of $\tilde{W}_n^{\rm spi}$ in Figure 3 (each histogram is based on $M = 2\,500$ values of these statistics).

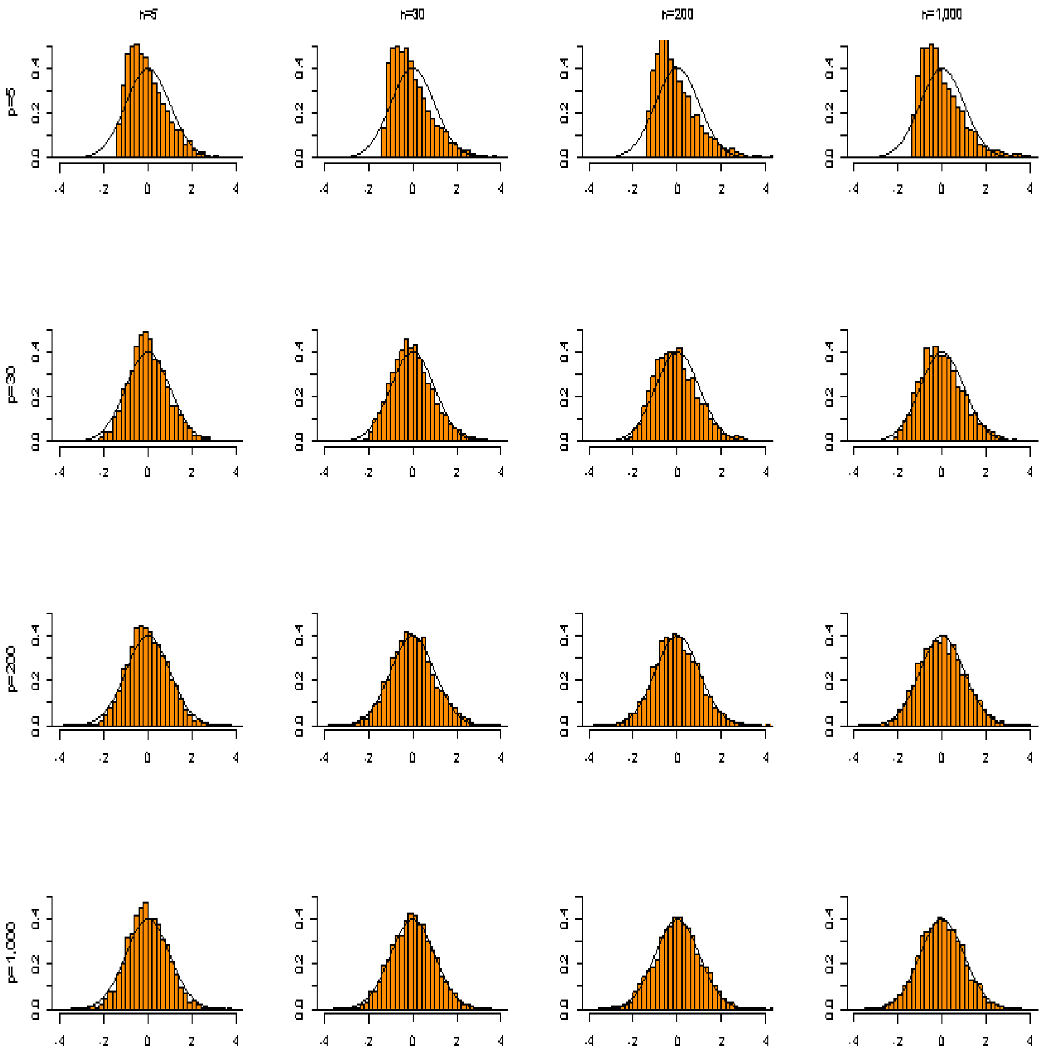

Figure 1: Histograms, for various values of n and p, of the standardized Watson statistic $\tilde{W}_n$ evaluated on $M = 2\,500$ random samples of size n from the p-dimensional FvML distribution with concentration $\kappa = 2$; see Section 5.1 for details.

From Theorem 3.1 and the discussion in Section 4, histograms are expected to be approximately standard normal as soon as $\min(n, p)$ is large, in a universal way (that is, irrespective of the relative sizes of n and p). Inspection of the results shows that, for all three setups, the standard normal approximation is valid for moderate-to-large values of n and p, irrespective of the value of p/n, which confirms our universal asymptotic results. Note also that, for small p and moderate-to-large n (that is, $p = 5$ and $n \geq 30$), histograms are approximately (standardized) chi-square, which is consistent with classical fixed-p asymptotic results; see Section 2.

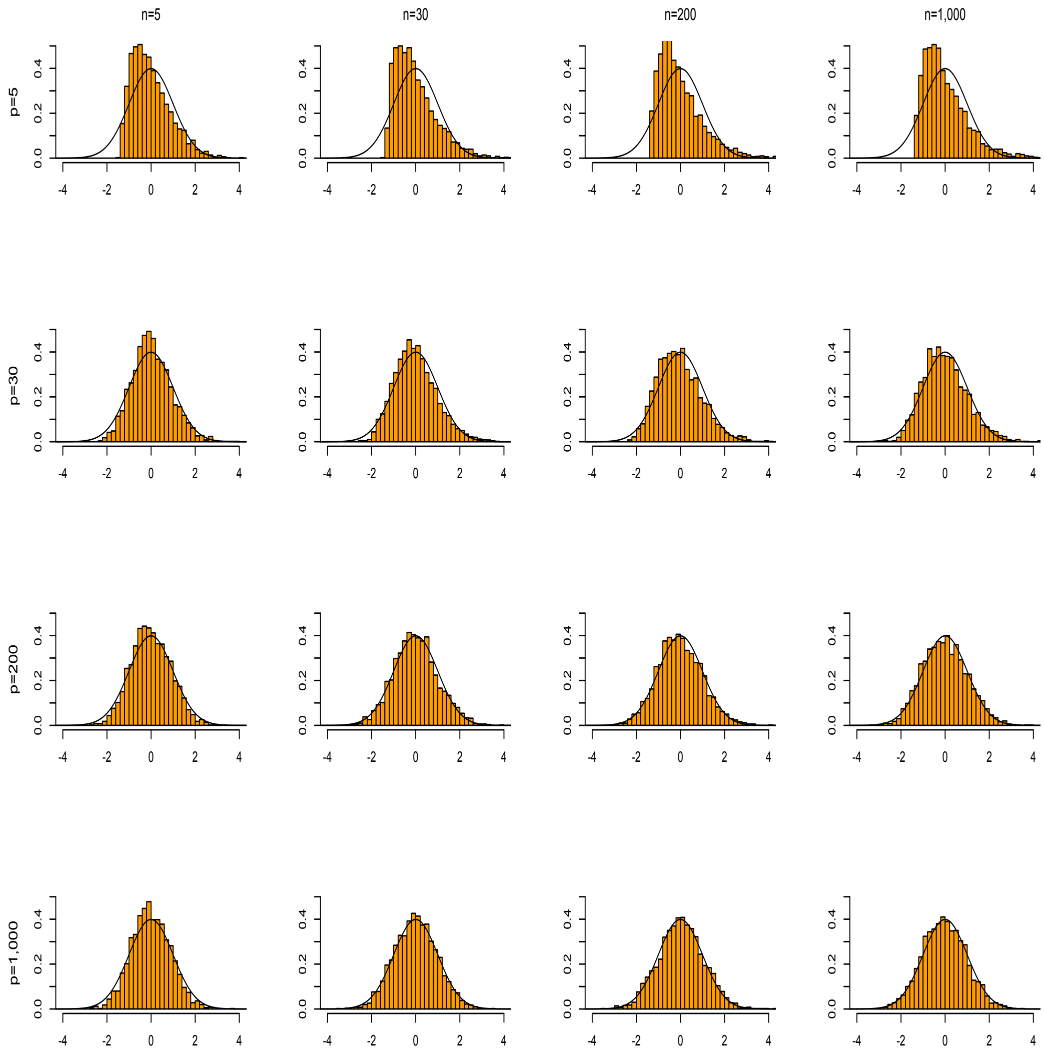

Figure 2: Histograms, for various values of n and p, of the standardized Watson statistic $\tilde{W}_n$ evaluated on $M = 2\,500$ random samples of size n from the p-dimensional Purkayastha distribution with concentration $\kappa = 1$; see Section 5.1 for details.

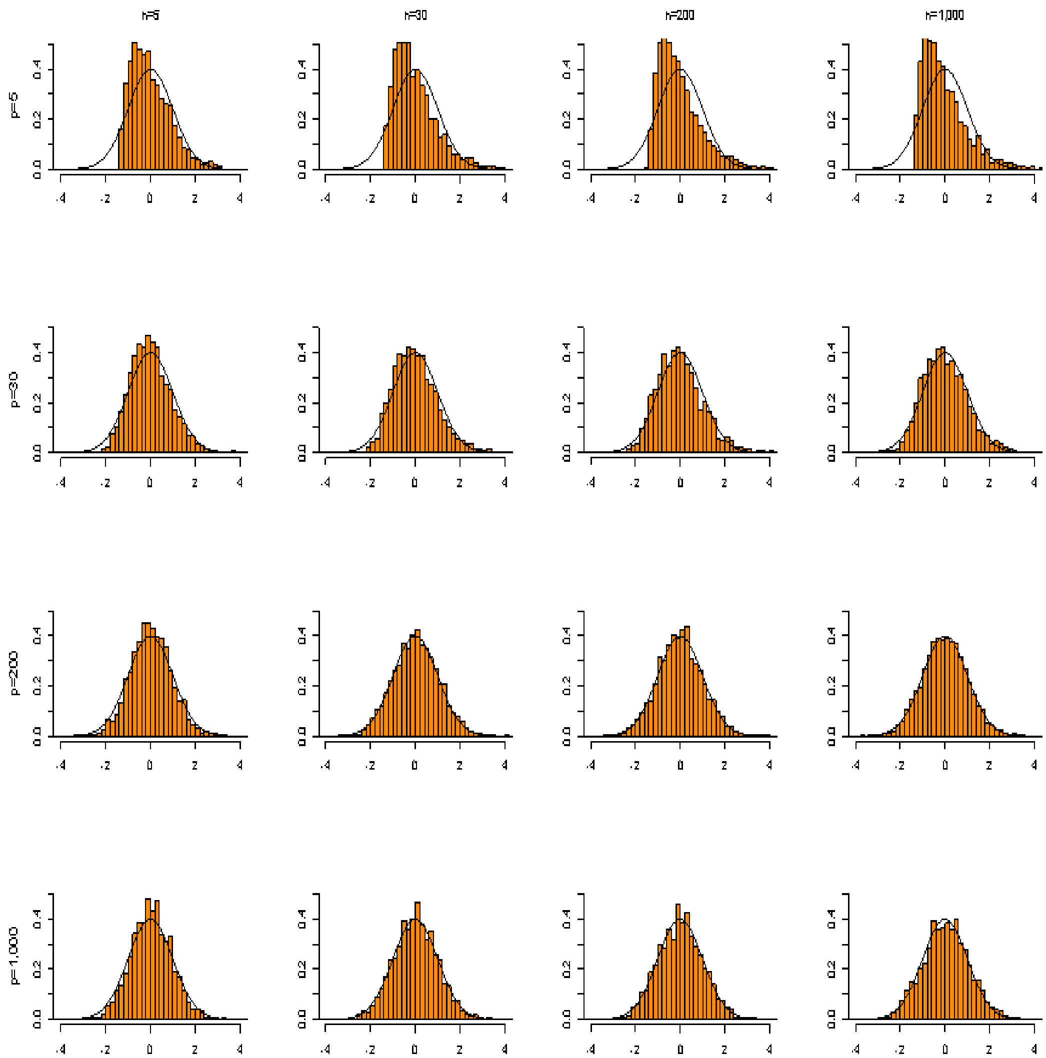

Figure 3: Histograms, for various values of n and p, of the test statistic $\tilde{W}_n^{\rm spi}$ for $\boldsymbol{\theta}_0$-spikedness evaluated on $M = 2\,500$ random samples of size n from the p-dimensional multinormal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}_p + (1/2)\boldsymbol{\theta}_0\boldsymbol{\theta}_0'$; see Section 5.1 for details.

To further assess the quality of the standard normal approximation at some relatively moderate dimensions p and sample sizes n, we conducted a second simulation, where we investigate how well the asymptotic Gaussian critical values approximate the (unknown) fixed-(n,p) corresponding quantiles of the Watson statistic under the null (this is of course of primary importance in the hypothesis context considered). To do so, for every $(n,p) \in C \times C$, with $C = \{10, 30, 100, 200\}$, we generated $M = 10\,000$ independent random samples from the FvML distributions $\mathcal{R}(\boldsymbol{\theta}_0, F_{p,1})$, where $\boldsymbol{\theta}_0$ is still the first vector of the canonical basis of $\mathbb{R}^p$. In line with the high-dimensional FvML distributions of Dryden (2005), we also conducted this simulation with the FvML distributions $\mathcal{R}(\boldsymbol{\theta}_0, F_{p,\sqrt{p}})$.

For every (n,p) and each concentration considered, we evaluated

$$\frac{1}{M} \sum_{i=1}^{M} \mathbb{I}[\tilde{W}_n > \Phi^{-1}(1-\alpha)]$$

($\mathbb{I}[A]$ stands for the indicator function of A), which is the empirical null size of the proposed high-dimensional Watson test. These rejection frequencies are reported in Table 1, which reveals that (i) the Gaussian approximation for $\tilde{W}_n$ indeed is reliable for relatively moderate values of n and p, and that (ii) the concentration does not have an important impact in practice.

		p			
	n	10	30	100	200
$\kappa = 1$	10	0.0622	0.0619	0.0634	0.0643
	30	0.0529	0.0591	0.0616	0.0647
	100	0.0523	0.0563	0.0540	0.0554
	200	0.0483	0.0517	0.0557	0.0537
$\kappa = \sqrt{k}$	10	0.0574	0.0630	0.0694	0.0655
	30	0.0550	0.0592	0.0645	0.0590
	100	0.0471	0.0545	0.0577	0.0565
	200	0.0478	0.0532	0.0582	0.0560

Table 1: For various values of n and p, null rejection frequencies of the high-dimensional Watson test computed from $M = 10\,000$ independent samples of size n generated according to the p-dimensional FvML distributions $\mathcal{R}(\boldsymbol{\theta}_0, F_{p,1})$ or $\mathcal{R}(\boldsymbol{\theta}_0, F_{p,\sqrt{p}})$; see Section 5.1 for details.

5.2. Behavior under the alternative

We conducted a last Monte-Carlo study to illustrate the non-null behavior of the proposed high-dimensional Watson test. To do so, we generated, for any $(n,p) \in C \times C$, with $C = \{20, 200, 1\,000\}$, independent random samples from the mixture-of-FvML distribution

$$\frac{1}{\ell} \mathcal{R}\big(\boldsymbol{\theta}_0, F_{p,\sqrt{p}}\big) + \Big(1 - \frac{1}{\ell}\Big) \mathcal{R}\big(\boldsymbol{\theta}_1, F_{p,\sqrt{p}}\big), \qquad \ell = 1, 2, 3, 4; \tag{5.6}$$

denoting by $e_{p,r}$ the rth vector of the canonical basis of $\mathbb{R}^p$, we took above $\boldsymbol{\theta}_0 = e_{p,1}$ and $\boldsymbol{\theta}_1 = (e_{p,1} + e_{p,p/4} - 2e_{p,p/2})/\sqrt{6}$. Clearly, $\ell = 1$ corresponds to the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ (FvML distribution with location $\boldsymbol{\theta}_0$), whereas $\ell = 2, 3, 4$ provide increasingly severe alternatives. For each (n,p)-value considered, Figure 4 reports the rejection frequencies of the high-dimensional Watson test based on $\tilde{W}_n$ (empirical rejection frequencies are based on $M = 2\,500$ replications). Clearly, this test exhibits non-trivial powers under the type of alternatives considered, irrespective of the value of (n,p).

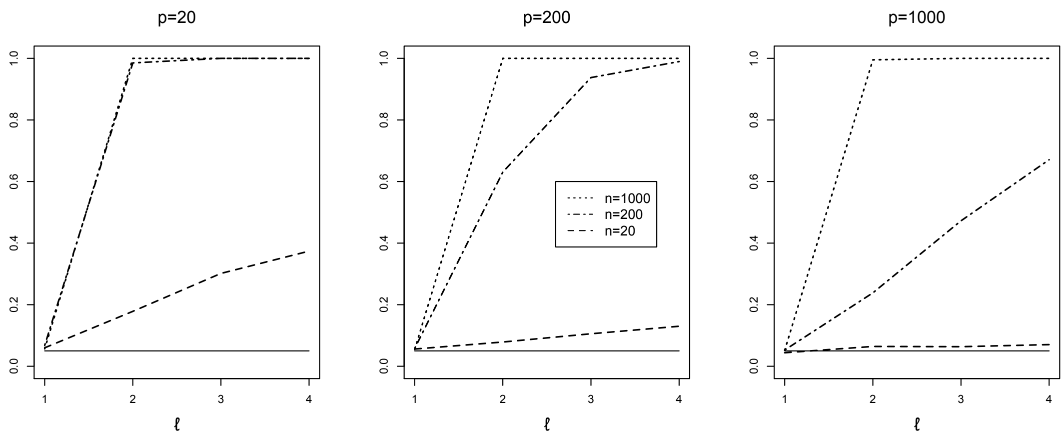

Figure 4: For various values of n and p, non-null rejection frequencies of the high-dimensional Watson test computed from $M = 2\,500$ independent samples of size n generated according to the p-dimensional mixture-of-FvML distributions in (5.6); see Section 5.2 for details.

Acknowledgements

Christophe Ley thanks the Fonds National de la Recherche Scientifique, Communauté Française de Belgique, for support via a Mandat de Chargé de Recherche. Davy Paindaveine's research is supported by an A.R.C. contract from the Communauté Française de Belgique and by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). All three authors would like to thank the Associate Editor and an anonymous referee for their comments that led to a significant improvement of the paper.

Appendix A. Proofs

We start with the proof of the main result, that is, Theorem 3.1. The proof will follow by applying the Slutsky Lemma to

$$\tilde{W}_n = \left(\frac{\sqrt{2(p_n - 1)}}{n\mathrm{E}[v_{n1}^2]} \sum_{1 \le i < j \le n} v_{ni} v_{nj} \mathbf{S}'_{ni} \mathbf{S}_{nj} \right) \Big/ \left(\frac{\frac{1}{n}\sum_{i=1}^n v_{ni}^2}{\mathrm{E}[v_{n1}^2]} \right) =: R_n / L_n. \tag{A.1}$$

The stochastic convergence of the denominator is taken care of in the following result.

Proposition A.1. *Under the assumptions of Theorem 3.1, $L_n \to 1$ in quadratic mean as $n \to \infty$.*

Proof of Proposition A.1. Since

$$\mathrm{E}\big[(L_n-1)^2\big] = \mathrm{E}\left[\left(\frac{\frac{1}{n}\sum_{i=1}^n v_{ni}^2}{\mathrm{E}[v_{n1}^2]}-1\right)^2\right] = \frac{1}{(\mathrm{E}[v_{n1}^2])^2}\,\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^n v_{ni}^2-\mathrm{E}[v_{n1}^2]\right)^2\right]$$

$$= \frac{1}{(\mathrm{E}[v_{n1}^2])^2}\,\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n v_{ni}^2\right] = \frac{\mathrm{Var}[v_{n1}^2]}{n(\mathrm{E}[v_{n1}^2])^2} \le \frac{\mathrm{E}[v_{n1}^4]}{n(\mathrm{E}[v_{n1}^2])^2}, \tag{A.2}$$

the result follows from Condition (iv) in Theorem 3.1. □

To establish Theorem 3.1, it is therefore sufficient to prove the following result.

Proposition A.2. *Under the assumptions of Theorem 3.1, R_n is asymptotically standard normal.*

The proof of this proposition is more delicate and will be based on the following martingale Central Limit Theorem; see Theorem 35.12 in Billingsley (1995).

Theorem A.1. *Assume that, for each n, $Z_{n1}, Z_{n2}, \ldots$ is a martingale relative to the filtration $\mathcal{F}_{n1}, \mathcal{F}_{n2}, \ldots$ and define $Y_{n\ell} = Z_{n\ell} - Z_{n,\ell-1}$. Suppose that the $Y_{n\ell}$'s have finite second-order moments and let $\sigma_{n\ell}^2 = \mathrm{E}[Y_{n\ell}^2 \,|\, \mathcal{F}_{n,\ell-1}]$ (with $\mathcal{F}_{n0} = \{\emptyset, \Omega\}$). Assume that $\sum_{\ell=1}^\infty Y_{n\ell}$ and $\sum_{\ell=1}^\infty \sigma_{n\ell}^2$ converge with probability 1. Then, if, for $n \to \infty$,*

$$\sum_{\ell=1}^\infty \sigma_{n\ell}^2 = \sigma^2 + o_{\mathrm{P}}(1), \tag{A.3}$$

where σ is a positive real number, and

$$\sum_{\ell=1}^\infty \mathrm{E}\big[Y_{n\ell}^2\,\mathbb{I}[|Y_{n\ell}| \ge \varepsilon]\big] \to 0 \quad \forall \varepsilon > 0, \tag{A.4}$$

we have that $\sigma^{-1}\sum_{\ell=1}^\infty Y_{n\ell}$ is asymptotically standard normal.

In order to apply this result, we need to identify the distinct quantities in the present setting. Let $\mathcal{F}_{n\ell}$ be the σ-algebra generated by $\mathbf{X}_{n1}, \ldots, \mathbf{X}_{n\ell}$ and denote by $\mathrm{E}_{n\ell}[.]$ the conditional expectation with respect to $\mathcal{F}_{n\ell}$. Then, letting

$$Y_{n\ell} := \mathrm{E}_{n\ell}[R_n] - \mathrm{E}_{n,\ell-1}[R_n] = \frac{\sqrt{2(p_n-1)}}{n\mathrm{E}[v_{n1}^2]}\sum_{i=1}^{\ell-1} v_{ni} v_{n\ell} \mathbf{S}_{ni}' \mathbf{S}_{n\ell}$$

for $\ell = 1, \ldots, n$ and (as in Billingsley, 1995) $Y_{n\ell} = 0$ for $\ell > n$, we clearly have that $R_n = \sum_{\ell=2}^n Y_{n\ell}$, where the $Y_{n\ell}$'s have finite second-order moments. Also, $\sum_{\ell=2}^\infty Y_{n\ell} = \sum_{\ell=2}^n Y_{n\ell}$ and $\sum_{\ell=2}^\infty \sigma_{n\ell}^2 = \sum_{\ell=2}^n \sigma_{n\ell}^2$, with $\sigma_{n\ell}^2 = \mathrm{E}_{n,\ell-1}[Y_{n\ell}^2]$ as in Theorem A.1, and both converge with probability 1, as required. Now, the crucial conditions (A.3) and (A.4) are shown to hold in the subsequent lemmas.

Lemma A.1. *Under the assumptions of Theorem 3.1, $\sum_{\ell=2}^n \sigma_{n\ell}^2 \to 1$ in quadratic mean as $n \to \infty$.*

Lemma A.2. *Under the assumptions of Theorem 3.1,* $\sum_{\ell=2}^{n} \mathrm{E}[Y_{n\ell}^2\, \mathbb{I}[|Y_{n\ell}| > \varepsilon]] \to 0$ *as* $n \to \infty$ *for any* $\varepsilon > 0$.

Before proving these lemmas, we recall that, under the assumptions of Theorem 3.1, the signs $\mathbf{S}_{ni}$ are uniformly distributed over $\mathcal{S}^{p_n-1}(\boldsymbol{\theta}_0^{\perp})$ (see Section 4) and that the v_{ni}'s are independent of the $\mathbf{S}_{ni}$'s, $i = 1, \ldots, n$. From Lemma A.1 in Paindaveine and Verdebout (2015a) it directly follows that, for fixed n, the quantities $\rho_{n,ij} := \mathbf{S}_{ni}'\mathbf{S}_{nj}$ are pairwise independent and satisfy $\mathrm{E}[\rho_{n,ij}] = 0$, $\mathrm{E}[\rho_{n,ij}^2] = 1/(p_n - 1)$, and $\mathrm{E}[\rho_{n,ij}^4] = 3/(p_n^2 - 1)$.

Proof of Lemma A.1. Rotational symmetry about $\boldsymbol{\theta}_0$ readily yields

$$\mathrm{E}[\mathbf{S}_{n\ell}\mathbf{S}_{n\ell}'] = \frac{1}{p_n - 1}\,(\mathbf{I}_{p_n} - \boldsymbol{\theta}_0\boldsymbol{\theta}_0').$$

The independence between the v_{ni}'s and $\mathbf{S}_{ni}$'s then provides

$$\begin{aligned}\sigma_{n\ell}^2 = \mathrm{E}_{n,\ell-1}[Y_{n\ell}^2] &= \frac{2(p_n-1)}{n^2(\mathrm{E}[v_{n1}^2])^2}\sum_{i,j=1}^{\ell-1} v_{ni}v_{nj}\mathrm{E}[v_{n\ell}^2]\mathbf{S}_{ni}'\mathrm{E}[\mathbf{S}_{n\ell}\mathbf{S}_{n\ell}']\mathbf{S}_{nj} \\ &= \frac{2}{n^2\mathrm{E}[v_{n1}^2]}\sum_{i,j=1}^{\ell-1} v_{ni}v_{nj}\rho_{n,ij}.\end{aligned}$$

Hence we obtain

$$\mathrm{E}\left[\sum_{\ell=2}^{n}\sigma_{n\ell}^2\right] = \frac{2}{n^2\mathrm{E}[v_{n1}^2]}\sum_{\ell=2}^{n}\sum_{i,j=1}^{\ell-1}\mathrm{E}[v_{ni}v_{nj}]\mathrm{E}[\rho_{n,ij}] = \frac{2}{n^2}\sum_{\ell=2}^{n}(\ell-1) = \frac{n-1}{n}. \tag{A.5}$$

Moreover, the pairwise independence of the $\rho_{n,ij}$'s entails

$$\mathrm{Var}\left[\sum_{\ell=2}^{n}\sigma_{n\ell}^2\right] = \frac{4}{n^4(\mathrm{E}[v_{n1}^2])^2}\mathrm{Var}\left[\sum_{\ell=2}^{n}\sum_{i,j=1}^{\ell-1} v_{ni}v_{nj}\rho_{n,ij}\right] = \frac{4}{n^4(\mathrm{E}[v_{n1}^2])^2}\left\{T_1^{(n)} + 4\,T_2^{(n)}\right\},$$

with

$$T_1^{(n)} := \mathrm{Var}\left[\sum_{\ell=2}^{n}\sum_{i=1}^{\ell-1} v_{ni}^2\right] = \mathrm{Var}\left[\sum_{i=1}^{n-1}(n-i)v_{ni}^2\right] = \sum_{i=1}^{n-1}(n-i)^2\,\mathrm{Var}[v_{n1}^2] \le n^3\,\mathrm{Var}[v_{n1}^2]$$

and

$$\begin{aligned}T_2^{(n)} &:= \mathrm{Var}\left[\sum_{\ell=2}^{n}\sum_{1\le i<j\le \ell-1} v_{ni}v_{nj}\rho_{n,ij}\right] = \mathrm{Var}\left[\sum_{1\le i<j\le n-1}(n-j)v_{ni}v_{nj}\rho_{n,ij}\right] \\ &= \sum_{1\le i<j\le n-1}(n-j)^2\mathrm{Var}[v_{ni}v_{nj}\rho_{n,ij}] = \sum_{1\le i<j\le n-1}(n-j)^2\mathrm{E}[v_{ni}^2u_{nj}^2\rho_{n,ij}^2] \\ &= \frac{(\mathrm{E}[v_{n1}^2])^2}{p_n-1}\sum_{1\le i<j\le n-1}(n-j)^2 \le \frac{n^4(\mathrm{E}[v_{n1}^2])^2}{p_n-1}.\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{Var}\left[\sum_{\ell=2}^{n}\sigma_{n\ell}^2\right] &\leq \frac{4\mathrm{Var}[v_{n1}^2]}{n(\mathrm{E}[v_{n1}^2])^2}+\frac{16}{p_n-1} \\ &\leq \frac{4\mathrm{E}[v_{n1}^4]}{n(\mathrm{E}[v_{n1}^2])^2}+\frac{16}{p_n-1} &\text{(A.6)}\\ &\to 0, &\text{(A.7)}\end{aligned}$$

in view of Conditions (ii) and (iv) from Theorem 3.1. Using (A.5) and (A.7) in

$$\mathrm{E}\left[\left(\sum_{\ell=2}^{n}\sigma_{n\ell}^2-1\right)^2\right]=\mathrm{Var}\left[\sum_{\ell=2}^{n}\sigma_{n\ell}^2\right]+\left(\mathrm{E}\left[\sum_{\ell=2}^{n}\sigma_{n\ell}^2-1\right]\right)^2$$

then establishes the result. □

Proof of Lemma A.2. Applying first the Cauchy-Schwarz inequality, then the Chebyshev inequality, yields

$$\sum_{\ell=2}^{n}\mathrm{E}[Y_{n\ell}^2\,\mathbb{I}[|Y_{n\ell}|>\varepsilon]]\leq\sum_{\ell=2}^{n}\sqrt{\mathrm{E}[Y_{n\ell}^4]}\,\sqrt{\mathrm{P}[|Y_{n\ell}|>\varepsilon]}\leq\frac{1}{\varepsilon}\sum_{\ell=2}^{n}\sqrt{\mathrm{E}[Y_{n\ell}^4]}\,\sqrt{\mathrm{Var}[Y_{n\ell}]}.$$

Noting that $\mathrm{Var}[Y_{n\ell}]\leq\mathrm{E}[Y_{n\ell}^2]=2(\ell-1)/n^2$, we obtain

$$\sum_{\ell=2}^{n}\mathrm{E}[Y_{n\ell}^2\,\mathbb{I}[|Y_{n\ell}|>\varepsilon]]\leq\frac{\sqrt{2}}{\varepsilon n}\sum_{\ell=2}^{n}\sqrt{\ell\,\mathrm{E}[Y_{n\ell}^4]}. \tag{A.8}$$

Using the fact that $0\leq v_{ni}\leq 1$ almost surely and the independence between the v_{ni}'s and the $\mathbf{S}_{ni}$'s, we get

$$\begin{aligned}\mathrm{E}\left[\left(\sum_{i=1}^{\ell-1}v_{ni}v_{n\ell}\rho_{n,i\ell}\right)^4\right] &= \sum_{i,j,r,s=1}^{\ell-1}\mathrm{E}\big[v_{n\ell}^4v_{ni}v_{nj}v_{nr}v_{ns}\rho_{n,i\ell}\rho_{n,j\ell}\rho_{n,r\ell}\rho_{n,s\ell}\big]\\ &= (\ell-1)(\mathrm{E}[v_{n1}^4])^2\mathrm{E}\big[\rho_{n,1\ell}^4\big]+3(\ell-1)(\ell-2)\mathrm{E}[v_{n1}^4](\mathrm{E}[v_{n1}^2])^2\mathrm{E}\big[\rho_{n,1\ell}^2\rho_{n,2\ell}^2\big]\\ &= \frac{3(\ell-1)}{p_n^2-1}(\mathrm{E}[v_{n1}^4])^2+\frac{3(\ell-1)(\ell-2)}{(p_n-1)^2}\mathrm{E}[v_{n1}^4](\mathrm{E}[v_{n1}^2])^2\\ &\leq \frac{3}{(p_n-1)^2}\Big[\ell(\mathrm{E}[v_{n1}^4])^2+\ell^2\mathrm{E}[v_{n1}^4](\mathrm{E}[v_{n1}^2])^2\Big],\end{aligned}$$

which yields

$$\begin{aligned}
\mathrm{E}\big[Y_{n\ell}^4\big] &\leq \frac{4(p_n-1)^2}{n^4(\mathrm{E}[v_{n1}^2])^4} \times \frac{3}{(p_n-1)^2}\Big[\ell(\mathrm{E}[v_{n1}^4])^2 + \ell^2\mathrm{E}[v_{n1}^4](\mathrm{E}[v_{n1}^2])^2\Big] \\
&\leq \frac{12}{n^4}\left[\ell\,\frac{(\mathrm{E}[v_{n1}^4])^2}{(\mathrm{E}[v_{n1}^2])^4} + \ell^2\,\frac{\mathrm{E}[v_{n1}^4]}{(\mathrm{E}[v_{n1}^2])^2}\right]. \qquad \text{(A.9)}
\end{aligned}$$

Plugging into (A.8), we conclude that

$$\begin{aligned}
\sum_{\ell=2}^n \mathrm{E}[Y_{n\ell}^2\,\mathbb{I}[|Y_{n\ell}|>\varepsilon]] &\leq \frac{\sqrt{24}}{\varepsilon n^3}\sum_{\ell=2}^n \sqrt{\ell^2\,\frac{(\mathrm{E}[v_{n1}^4])^2}{(\mathrm{E}[v_{n1}^2])^4} + \ell^3\,\frac{\mathrm{E}[v_{n1}^4]}{(\mathrm{E}[v_{n1}^2])^2}} \\
&\leq \frac{\sqrt{24}}{\varepsilon n^3}\sum_{\ell=2}^n \left(\ell\,\frac{\mathrm{E}[v_{n1}^4]}{(\mathrm{E}[v_{n1}^2])^2} + \ell^{3/2}\sqrt{\frac{\mathrm{E}[v_{n1}^4]}{(\mathrm{E}[v_{n1}^2])^2}}\right) \\
&\leq O(n^{-1})\,\frac{\mathrm{E}[v_{n1}^4]}{(\mathrm{E}[v_{n1}^2])^2} + O(n^{-1/2})\sqrt{\frac{\mathrm{E}[v_{n1}^4]}{(\mathrm{E}[v_{n1}^2])^2}},
\end{aligned}$$

which, in view of Condition (iv) from Theorem 3.1, is indeed $o(1)$. □

It remains to prove Theorem 3.2 and Theorem 3.3.

Proof of Theorem 3.2. In this proof, C will stand for a generic constant that may change from line to line. Applying (with $\delta=1$) the theorem in Heyde and Brown (1970) to the martingale $R_n = \sum_{\ell=2}^n Y_{n\ell}$ considered in the previous proof readily provides

$$\sup_{z\in\mathbb{R}}\big|\mathrm{P}[R_n \leq s_n z] - \Phi(z)\big| \leq C\bigg(\sum_{\ell=2}^n \mathrm{E}\big[Y_{n\ell}^4\big] + \mathrm{Var}\bigg[\sum_{\ell=2}^n \sigma_{n\ell}^2\bigg]\bigg)^{1/5},$$

with $s_n^2 = \sum_{\ell=2}^n \mathrm{E}\big[\sigma_{n\ell}^2\big] = (n-1)/n$ (see (A.5)). Using (A.6) and (A.9) then yields that, for n large enough,

$$\sup_{z\in\mathbb{R}}\big|\mathrm{P}[R_n \leq s_n z] - \Phi(z)\big| \leq C\big(c_n + c_n^2 + p_n^{-1}\big)^{1/5} \leq C\big(c_n + p_n^{-1}\big)^{1/5}, \qquad \text{(A.10)}$$

where we let

$$c_n := \frac{\mathrm{E}[v_{n1}^4]}{n(\mathrm{E}[v_{n1}^2])^2}\,.$$

It therefore only remains to show that, for n large enough,

$$\sup_{z\in\mathbb{R}}\big|\mathrm{P}[\tilde{W}_n \leq s_n z] - \mathrm{P}[R_n \leq s_n z]\big| \leq C\big(c_n + p_n^{-1}\big)^{1/5}. \qquad \text{(A.11)}$$

To do so, recall (A.1) and write

$$
\begin{aligned}
&\big|\mathrm{P}[\tilde{W}_n \le z] - \mathrm{P}[R_n \le z]\big| = \big|\mathrm{P}[R_n \le L_n z] - \mathrm{P}[R_n \le z]\big| \\
&\qquad = \mathrm{P}[\min(z, L_n z) \le R_n \le \max(z, L_n z)] \\
&\qquad \le \mathrm{P}[|L_n - 1| > c_n^{2/5}] + \mathrm{P}[\min(z, L_n z) \le R_n \le \max(z, L_n z), |L_n - 1| \le c_n^{2/5}] \\
&\qquad =: E_n + F_n,
\end{aligned}
$$

say. Using the Markov inequality and (A.2), we readily obtain

$$
E_n \le \frac{\mathrm{E}\big[|L_n - 1|^2\big]}{c_n^{4/5}} \le \frac{c_n}{c_n^{4/5}} = c_n^{1/5}.
$$

As for F_n, applying (A.10) to

$$
\begin{aligned}
F_n &\le \mathrm{P}[\min((1 \pm c_n^{2/5})z) \le R_n \le \max((1 \pm c_n^{2/5})z)] \\
&\le \mathrm{P}[R_n \le \max((1 \pm c_n^{2/5})z)] - \mathrm{P}[R_n \le \min((1 \pm c_n^{2/5})z)]
\end{aligned}
$$

yields

$$
\begin{aligned}
F_n &\le C\big(c_n + p_n^{-1}\big)^{1/5} + \Phi(\max((1 \pm c_n^{2/5})z)/s_n) - \Phi(\min((1 \pm c_n^{2/5})z)/s_n) \\
&\le C\big(c_n + p_n^{-1}\big)^{1/5} + \frac{2c_n^{2/5}|z|}{s_n\sqrt{2\pi}} \exp\Big(- \frac{(1 + \xi_{n,z} c_n^{2/5})^2 z^2}{2s_n^2}\Big)
\end{aligned}
$$

for some $\xi_{n,z} \in (-1, 1)$. For n large enough, we therefore have

$$
F_n \le C\big(c_n + p_n^{-1}\big)^{1/5} + \frac{2c_n^{2/5}|z|}{s_n\sqrt{2\pi}} \exp\Big(- \frac{(1/2)^2 z^2}{2s_n^2}\Big) \le C\big(c_n + p_n^{-1}\big)^{1/5} + Cc_n^{1/5},
$$

so that

$$
E_n + F_n \le C\left(c_n + p_n^{-1}\right)^{1/5} + Cc_n^{1/5} \le C\left(c_n + p_n^{-1}\right)^{1/5}.
$$

We conclude that, still for n large enough,

$$
\sup_{z\in\mathbb{R}} \big|\mathrm{P}[\tilde{W}_n \le s_n z] - \mathrm{P}[R_n \le s_n z]\big| = \sup_{z\in\mathbb{R}} \big|\mathrm{P}[\tilde{W}_n \le z] - \mathrm{P}[R_n \le z]\big| \le C\big(c_n + p_n^{-1}\big)^{1/5},
$$

which is (A.11). This establishes the result. □

Proof of Theorem 3.3. Decompose $\tilde{W}_n - \tilde{Q}_n$ into $A_n + B_n$, with

$$
A_n = \left(\frac{\mathrm{E}[v_{n1}^2]}{\frac{1}{n}\sum_{i=1}^n v_{ni}^2} - 1\right) \frac{\sqrt{2(p_n - 1)}}{n\mathrm{E}[v_{n1}^2]} \sum_{1\le i<j\le n} v_{ni} v_{nj} \mathbf{S}_{ni}' \mathbf{S}_{nj}
$$

and

$$B_n = \frac{\sqrt{2(p_n - 1)}}{n} \sum_{1 \le i < j \le n} \left(\frac{v_{ni} v_{nj}}{\mathrm{E}[v_{n1}^2]} - 1 \right) \mathbf{S}'_{ni} \mathbf{S}_{nj}.$$

Propositions A.1 and A.2 readily entail that $A_n = o_{\mathrm{P}}(1)$ as $n \to \infty$. As for B_n, we have (see the beginning of the Appendix for a recall on some results regarding expectations of the signs $\mathbf{S}_{ni}$)

$$\mathrm{E}[B_n^2] = \frac{2(p_n - 1)}{n^2} \sum_{1 \le i < j \le n} \mathrm{E}\left[\left(\frac{v_{ni} v_{nj}}{\mathrm{E}[v_{n1}^2]} - 1 \right)^2 (\mathbf{S}'_{ni} \mathbf{S}_{nj})^2 \right] = \frac{2}{n^2} \sum_{1 \le i < j \le n} \mathrm{E}\left[\left(\frac{v_{ni} v_{nj}}{\mathrm{E}[v_{n1}^2]} - 1 \right)^2 \right]$$

$$= \frac{n-1}{n} \mathrm{E}\left[\left(\frac{v_{n1} v_{n2}}{\mathrm{E}[v_{n1}^2]} - 1 \right)^2 \right] = \frac{2(n-1)}{n} \mathrm{E}\left[1 - \frac{v_{n1} v_{n2}}{\mathrm{E}[v_{n1}^2]} \right] = \frac{2(n-1)}{n} \left(1 - \frac{(\mathrm{E}[v_{n1}])^2}{\mathrm{E}[v_{n1}^2]} \right),$$

which, in view of Condition (v), is $o(1)$ as $n \to \infty$. The result follows. □

Banerjee, A., Dhillon, I., Ghosh, J., Sra, S., 2003. Generative model-based clustering of directional data. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 19–28.

Banerjee, A., Dhillon, I., Ghosh, J., Sra, S., 2005. Clustering on the unit hypersphere using von mises-fisher distributions. J. Mach. Learn. Res. 6, 1345–1382.

Banerjee, A., Ghosh, J., 2002. Frequency sensitive competitive learning for clustering on high-dimensional hyperspheres. In Proceedings International Joint Conference on Neural Networks, 1590–1595.

Banerjee, A., Ghosh, J., 2004. Frequency sensitive competitive learning for scalable balanced clustering on high-dimensional hyperspheres. IEEE T. Neural Networ. 15, 702–719.

Baricz, A., Ponnusamy, S., 2013. On turán type inequalities for modified bessel functions. Proc. Amer. Math. Soc. 141 (523–532).

Billingsley, P., 1995. Probability and Measure, 3rd Edition. Wiley, New York, Chichester.

Cai, T., Fan, J., Jiang, T., 2013. Distributions of angles in random packing on spheres. J. Mach. Learn. Res. 14, 1837–1864.

Cai, T., Jiang, T., 2012. Phase transition in limiting distributions of coherence of high-dimensional random matrices. J. Multivariate Anal. 107, 24–39.

Chen, S., Qin, Y., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Statist. 38, 808–835.

Chen, S. X., Zhang, L.-X., Zhong, P.-S., 2010. Tests for high-dimensional covariance matrices. J. Amer. Statist. Assoc. 105, 810–819.

Dryden, I. L., 2005. Statistical analysis on high-dimensional spheres and shape spaces. Ann. Statist. 33, 1643–1665.

Heyde, C. C., Brown, B. M., 1970. On the departure from normality of a certain class of martingales. Ann. Mathem. Statist. 41, 2161–2165.

Jiang, T., Yang, F., 2013. Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. Ann. Statist. 41, 2029–2074.

Johnstone, I. M., 2001. On the distribution of the largest eigenvalue in principal components analysis. Ann. Statist. 29, 295–327.

Ledoit, O., Wolf, M., 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann. Statist. 30, 1081–1102.

Li, J., Chen, S. X., 2012. Two sample tests for high-dimensional covariance matrices. Ann. Statist. 40, 908–940.

Mardia, K. V., Jupp, P. E., 2000. Directional Statistics. John Wiley & Sons.

Onatski, A., Moreira, M., Hallin, M., 2013. Asymptotic power of sphericity tests for high-dimensional data. Ann. Statist. 41, 1204–1231.

Paindaveine, D., Verdebout, T., 2015a. On high-dimensional sign tests. Bernoulli, to appear.

Paindaveine, D., Verdebout, T., 2015b. Optimal rank-based tests for the location parameter of a rotationally symmetric distribution on the hypersphere. In: M. Hallin, D. Mason, D. Pfeifer, J. Steinebach (Eds.), Mathematical Statistics and Limit Theorems: Festschrift in Honor of Paul Deheuvels. Springer, to appear.

Saw, J. G., 1978. A family of distributions on the m-sphere and some hypothesis tests. Biometrika 65, 69–73.

Schott, J., 2007. Some high-dimensional tests for a one-way manova. J. Multivariate Anal. 98, 1825–1839.

Srivastava, M. S., Fujikoshi, Y., 2006. Multivariate analysis of variance with fewer observations than the dimension. J. Multivariate Anal. 97, 1927–1940.

Srivastava, M. S., Katayama, S., Kano, Y., 2013. A two sample test in high dimensional data. J. Multivariate Anal. 114, 349–358.

Srivastava, M. S., Kubokawa, T., 2013. Tests for multivariate analysis of variance in high dimension under non-normality. J. Multivariate Anal. 115, 204–216.

Stam, A. J., 1982. Limit theorems for uniform distributions on spheres in high-dimensional euclidean spaces. J. Appl. Probab. 19, 221–228.

Watson, G., 1944. A Treatise on the Theory of Bessel Functions, second edition Edition. Cambridge University Press.

Watson, G. S., 1983a. Limit theorems on high-dimensional spheres and stiefel manifolds. In: Karlin, S., Amemiya, T., Goodman, L. A. (Eds.), Studies in Econometrics, Time Series, and Multivariate Statistics. Academic Press, 559–570, New York, pp. 559–570.

Watson, G. S., 1983b. Statistics on Spheres. Wiley, New York.

Watson, G. S., 1988. The langevin distribution on high dimensional spheres. J. Appl. Statist. 15, 123–130.