Submitted to Bernoulli

Inference on location for noisy directional data: A Le Cam approach to quantify the value of the hyperspherical a priori information

DIEGO BOLÓN[a], DAVY PAINDAVEINE[b] and THOMAS VERDEBOUT[c]

CITMAga, Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, [a]*diego.bolon.rodriguez@usc.es*

ECARES and Department of Mathematics, Université libre de Bruxelles, [b]*Davy.Paindaveine@ulb.be,* [c]*Thomas.Verdebout@ulb.be*

We consider models for noisy directional data, in which a radial noise with magnitude σ^2 makes the observations deviate from their theoretical hyperspherical sample space, namely a hypersphere centered at θ and with radius r. We consider inference — hypothesis testing, point estimation, and confidence zone estimation — on the location parameter θ, in a framework where both r and σ^2 remain unspecified. We introduce several asymptotic scenarios in which the radius of the hypersphere and, most importantly, the noise magnitude may depend on the sample size n in an essentially arbitrary way. This allows us to consider very diverse cases, in which the a priori information that the data belong to a hypersphere is more and more, or on the contrary less and less, relevant. We base our investigation on Le Cam's asymptotic theory of statistical experiments and aim at a full understanding of the resulting limiting experiments. The corresponding contiguity rates, that characterize how easy/hard inference on θ is, reveal rather counter-intuitive results in some scenarios. We build locally asymptotically optimal tests and estimators, that turn out to be adaptively optimal across all asymptotic scenarios. We show that, in standard asymptotic scenarios, classical procedures that would ignore the hyperspherical a priori information are rate-consistent but do not achieve efficiency bounds, and that, in non-standard asymptotic scenarios, such classical procedures are not even rate-consistent. We investigate the finite-sample relevance of our results through Monte Carlo exercises.

Keywords: Contiguity rates; directional statistics; double asymptotics; Le Cam's theory of asymptotic experiments; multivariate location; one-step estimation; strong identifiability

1. Introduction

Multivariate location problems are probably among the most extensively studied problems in the statistical literature. The most standard tests for the one-sample and two-sample location problems are certainly the Hotelling tests; see Hotelling (1931). They are still studied a lot nowadays; we refer, e.g., to Chen et al. (2011) and Feng et al. (2017) for regularized one-sample Hotelling tests, and to Li et al. (2020) for an adaptable two-sample Hotelling test. Hotelling tests are based on sample means, hence are not robust to outliers or heavy-tailed distributions. This motivated the introduction of nonparametric tests, and in particular of sign and signed-rank tests, for multivariate location problems; see, among many others, Randles (2000), Hallin and Paindaveine (2002), Larocque, Nevalainen and Oja (2007), Wang, Peng and Li (2015) and Feng, Zou and Wang (2016). Recent contributions related to inference for multivariate location include Agostinelli and Greco (2019) that studied weighted likelihood estimation, Frahm, Nordhausen and Oja (2020), that proposed location estimators for incomplete and dependent multivariate data, Dürre and Paindaveine (2022), that defined simplex-based affine-equivariant location estimators, Chakraborty and Chaudhuri (2017), Kock and Preinerstorfer (2019) and Kock and Preinerstorfer (2023), that tackled the high-dimensional case, and Ley et al. (2013), Paindaveine and Verdebout (2017), Paindaveine and Verdebout (2020a) and Paindaveine and Verdebout (2020b), that considered location problems in a directional data framework.

Now, be it in low- or high-dimensional statistics, it is more and more common to assume that the actual dimensionality of the data, k say, is smaller than the dimension d of the ambient space; in high dimensions, see, e.g., Wright and Ma (2022). Under a "linearity" assumption stating that the data arise from a noisy version of a distribution concentrated on a k-dimensional hyperplane, the classical methodology then relies on PCA, but of course, it is more promising and general to consider nonlinear dimension reduction techniques allowing the data to deviate from a non-flat, k-dimensional, manifold. In practice, both the "intrinsic" dimension k and the corresponding manifold are often unspecified, and a vast literature has tackled the problem of recovering these key unknown quantities; we refer, e.g., to Donoho and Grimes (2003), Maggioni, Minsker and Strawn (2016), and the references therein. In particular, recovering spheres plays an important role in various scientific domains. Recent works studying statistical inference for such data mildly deviating from a manifold include for instance Shapiro, Xie and Zhang (2021) and Cheng and Xie (2024), that consider goodness-of-fit testing and two-sample testing, respectively.

Typically, results in the literature provide consistency rates and, at best, establish their minimax optimality under suitable assumptions; see, e.g., Shapiro, Xie and Zhang (2021) and Cheng and Xie (2024). To the best of our knowledge, finer optimality/efficiency results, that would for instance state optimality in the Le Cam sense, have never been considered in this framework. The present work aims at deriving such stronger results. Since there is no free lunch, the price to pay to achieve this is to make more assumptions on the shape of the underlying manifold, e.g., to assume that this manifold is a hypersphere or a hypertorus. In this paper, we indeed follow this path and adopt a framework in which the sample at hand is made of noisy versions of random vectors taking their values on a hypersphere of $\mathbb{R}^d$. This assumption will allow us to provide strong optimality results as the ones stated above without severely compromising the practical applicability of the proposed model. In fact, similar frameworks where sphere detection/estimation take an important role have recently been considered in medical imaging (van der Glas et al., 2002) and computer vision (Sandoval et al., 2020).

More precisely, the distributional framework we will consider is described by triangular arrays of observations X_{ni}, $i = 1, \ldots, n$, $n = 1, 2, \ldots$, whose distribution will be characterized by a real d-vector θ and two sequences of positive real numbers (r_n) and (σ_n^2). We will throughout denote as $\mathrm{P}^n_{\theta, r_n, \sigma_n^2}$ the hypothesis under which the random d-vectors X_{ni}, $i = 1, \ldots, n$ are given by

$$X_{ni} = \theta + r_n e^{\varepsilon_{ni}} U_{ni}, \tag{1.1}$$

where the ε_{ni}'s form a random sample from the Gaussian distribution with mean zero and variance σ_n^2, the U_{ni}'s form a random sample from the uniform distribution over the unit sphere $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ of $\mathbb{R}^d$, and the ε_{ni}'s and U_{ni}'s are mutually independent. The observations X_{ni}, $i = 1, \ldots, n$, can thus be regarded as a noisy version of a sample that would concentrate on a manifold, namely the hypersphere with centre θ and radius r_n, with roughly as many observations inside and outside the hypersphere (as $e^{\varepsilon_{n1}}$ has median one). While the shape of the underlying manifold is fixed by the model, the manifold itself remains unspecified, as both θ and r_n are unknown parameters. A key parameter is of course σ_n^2, that characterizes the magnitude of the noise. For small σ_n^2, observations are close to the underlying manifold, and the larger σ_n^2, the more observations will tend to deviate from it. In this paper, we will consider inference on the centre θ of the underlying hypersphere.

Before proceeding, we briefly describe two real-data settings in which the model (1.1) is natural and the parameter of interest is the centre θ. *(A) Earth surface points in ECEF*. Let $X_{ni} \in \mathbb{R}^3$ denote *Earth-Centered Earth-Fixed* coordinates of a globally distributed set of surface sites (e.g., GNSS stations or cities with elevations). These points concentrate near a sphere centered at the geocentre θ, while topography and geoid undulations induce mostly radial departures. A convenient description is precisely (1.1), with U_{ni} approximately uniform on $\mathcal{S}^2$ under global coverage and ε_{ni} modeling multiplicative radial noise that summarizes elevation variability. *(B) LEO satellite shell at a snapshot*. Consider a single near-circular orbital shell observed at one epoch t_0, with $X_{ni} \in \mathbb{R}^3$ the satellites' ECEF positions $X_{ni}(t_0)$. For a given shell, satellites lie close to a sphere centered at θ; small eccentricities, short-term dynamics over the snapshot, and ephemeris errors contribute mainly radial dispersion captured by the ε_{ni}'s, while the instantaneous directions U_{ni} are broadly spread on $\mathcal{S}^2$. In both (A)–(B), inference on θ is to be considered in a spherical framework that offers a small radial noise and directions that are essentially uniformly distributed.

In the context of the model (1.1), it should be clear that the sequence (σ_n^2) will play a key role when making inference on the centre θ of the underlying hypersphere: if (σ_n^2) converges to zero, then observations will tend to be closer and closer to the manifold, which should make inference on θ easier; would all observations be on a common hyperpshere with centre θ, the *exact* value of θ could indeed be obtained almost surely from a random sample of size $n = d + 1$. When $(\sigma_n^2) \to 0$, it may thus be expected that faster-than-$\sqrt{n}$ consistency rates will result from this "strong identifiability" of θ; see Paindaveine and Verdebout (2020b) for a similar phenomenon in spherical location problems. On the contrary, a sequence (σ_n^2) that diverges to infinity should intuitively make inference on θ harder, which should translate into slower-than-$\sqrt{n}$ consistency rates. To explore this, we will actually consider three different asymptotic scenarios: (i) the favourable scenario in which $(\sigma_n^2) \to 0$ where super-efficiency is expected, (ii) a neutral, standard, scenario in which $(\sigma_n^2) \to \sigma^2 (> 0)$, and (iii) the unfavourable scenario in which $(\sigma_n^2) \to \infty$.

Our objectives in this framework are four-fold: first, we want to quantify in a precise way how much it helps, in each of these three scenarios, to have the a priori information that observations are noisy hyperspherical data. We aim at doing so not only in terms of consistency rates but also by comparing the corresponding efficiency bounds. Second, we aim at defining statistical procedures that will be asymptotically optimal across the three asymptotic scenarios above, in an adaptive way (that is, the proposed procedures should achieve asymptotic optimality in each scenario without using knowledge of which scenario is the "true" one). Third, we want to characterize the possible cost of unspecification of r_n and σ_n^2, that is, we want to answer the following questions: has the unspecification of the radius of the underlying hypersphere an asymptotic cost? Is there a cost of not knowing the noise magnitude? Fourth and last, our goal is to compare the performances of the proposed optimal procedures with those of procedures that would ignore the hyperspherical a priori information such as Hotelling tests. In particular, we will investigate whether this cost is in terms of consistency rate or whether it is only in terms of asymptotic variance/power at a common rate.

To fulfill the above objectives, we will adopt an approach that relies on Le Cam's theory of asymptotic experiments (see, e.g., Le Cam and Yang, 2000 or van der Vaart, 1998). Due to the particularities of the model (1.1), the Le Cam's theory is developed here in a non-standard setting involving triangular arrays of observations in order to deal with the three aforementioned scenarios. In each one of them, we will determine the limiting experiment, hence also identify the corresponding contiguity rate. These contiguity rates bring an important information since they contribute to characterizing how easy/hard it is to perform inference on θ in each scenario. Most importantly, the Le Cam approach will pave the way to defining asymptotically optimal procedures in the various scenarios and to quantifying their asymptotic performance. A key point will of course be to control the estimation of the underlying nuisance parameters, namely r_n and σ_n^2, and to investigate adaptivity of the proposed procedures in the underlying asymptotic scenarios. We will consider all inference problems on θ: hypothesis testing, point estimation and confidence zone estimation. While estimation, as often when a Le Cam approach is adopted, will rely on a one-step methodology, the construction of suitable, rate-consistent, preliminary estimators will be challenging in non-standard asymptotic scenarios. We will also consider hypothesis testing in a two-sample version of the above model, in a general framework that allows both populations to be in different asymptotic scenarios.

The outline of the paper is as follows. In Section 2, we determine the limiting experiments that are associated with the three aforementioned scenarios, which will provide in particular the corresponding contiguity rates. In Section 3, we consider hypothesis testing on θ. In Section 3.1, we construct the proposed test and study its asymptotic null behavior. Then, in Section 3.2, we investigate its asymptotic behavior under contiguous alternatives and compare its performance to those of two classical tests that ignore the hyperspherical a priori information. In Section 4, we turn to point estimation and confidence zone estimation of θ. In Section 5, we conduct Monte Carlo exercises to study the finite-sample perfomance of the proposed tests (Section 5.1) and estimators (Section 5.2). Finally, in Section 6, we provide a brief wrap up and we discuss perspectives for future research. All proofs are deferred to the supplement Bolón, Paindaveine and Verdebout (2025).

Before proceeding, we collect here some notation. Equality in distribution and convergence in distribution will be denoted as $=_{\mathcal{D}}$ and $\to_{\mathcal{D}}$, respectively. The notation ":=" will be used to define objects. Throughout, A' will be the transpose matrix of A, and $\|x\| := \sqrt{x'x}$ the Euclidean norm of the vector x (as already used above). The notation χ^2_d will denote the chi-square distribution with d degrees of freedom and $\chi^2_{d,1-\alpha}$ will stand for the corresponding upper α-quantile. Both e^z and $\exp(z)$ will refer to the exponential of the real number z. Unless otherwise mentioned, all deterministic and stochastic convergences will be as the sample size n diverges to infinity. This includes o and O statements, as well as o_{P} and O_{P} ones.

2. Limiting experiments

Since σ_n^2 is assumed to be positive, the distribution of the random vector X_{ni} in (1.1) is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. It is easy to check that the corresponding density is given by

$$f_{\theta,r_n,\sigma_n^2}(x) = \frac{\|x-\theta\|^{-d}}{\omega_d\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(\log\|x-\theta\| - \log r_n)^2}{2\sigma_n^2}\right),$$

where ω_d is the surface area measure of the unit sphere $\mathcal{S}^{d-1}$. Consider then local log-likelihood ratios of the form

$$\Lambda^n_{\theta+\nu_n\tau/\theta;r_n,\sigma_n^2} := \log\left(\frac{d\mathrm{P}^n_{\theta+\nu_n\tau,r_n,\sigma_n^2}}{d\mathrm{P}^n_{\theta,r_n,\sigma_n^2}}\right) = \log\left(\frac{\prod_{i=1}^n f_{\theta+\nu_n\tau,r_n,\sigma_n^2}(X_{ni})}{\prod_{i=1}^n f_{\theta,r_n,\sigma_n^2}(X_{ni})}\right)$$

$$= -d\sum_{i=1}^n \big(\log\|X_{ni}-\theta-\nu_n\tau\| - \log\|X_{ni}-\theta\|\big)$$

$$-\frac{1}{2\sigma_n^2}\sum_{i=1}^n \Big\{(\log\|X_{ni}-\theta-\nu_n\tau\|-\log r_n)^2-(\log\|X_{ni}-\theta\|-\log r_n)^2\Big\}, \tag{2.1}$$

where θ and τ are fixed d-vectors, and (ν_n) is a positive sequence converging to zero at a suitable rate. The following result determines the corresponding limiting experiments in the three scenarios described in the introduction (the proof, which is long and technical, is provided in Section S.2)[1].

Theorem 2.1. *Fix d-vectors θ and τ, and positive real sequences (r_n) and (σ_n^2). Then, irrespective of whether (i) (σ_n^2) converges to zero, (ii) (σ_n^2) converges to $\sigma^2 \in (0,\infty)$, or (iii) (σ_n^2) diverges to infinity, there exist a positive real sequence (ν_n) and a $d \times d$ non-singular matrix Γ such that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,*

$$\Lambda^n_{\theta+\nu_n\tau/\theta;r_n,\sigma_n^2} = \tau'\Delta_n(\theta) - \frac{1}{2}\tau'\Gamma\tau + o_{\mathrm{P}}(1), \tag{2.2}$$

where, still under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$, the central sequence

$$\Delta_n(\theta) = \nu_n \sum_{i=1}^n \left(d + \frac{\log\|Z_{ni}(\theta)\| - \log r_n}{\sigma_n^2}\right)\frac{Z_{ni}(\theta)}{\|Z_{ni}(\theta)\|^2}, \quad \textit{with } Z_{ni}(\theta) := X_{ni} - \theta,$$

is asymptotically normal with mean zero and covariance matrix Γ. (i) If $(\sigma_n^2) \to 0$, then this holds with $\nu_n = r_n\sigma_n/\sqrt{n}$ and $\Gamma = (1/d)I_d$; (ii) if $(\sigma_n^2) \to \sigma^2 \in (0,\infty)$, then this holds with $\nu_n = r_n/\sqrt{n}$ and

$$\Gamma = \frac{\exp(2\sigma^2)}{d}\left((d-2)^2 + \frac{1}{\sigma^2}\right)I_d; \tag{2.3}$$

(iii) if $(\sigma_n^2) \to \infty$ in such a way that $\exp((4+\varepsilon)\sigma_n^2) = o(n)$ for some $\varepsilon > 0$, then, this holds with $\nu_n = r_n\sigma_n/(\sqrt{n}\exp(\sigma_n^2))$ and $\Gamma = (1/d)I_d$ for $d = 2$, and with $\nu_n = r_n/(\sqrt{n}\exp(\sigma_n^2))$ and $\Gamma = ((d-2)^2/d)I_d$ for $d \geq 3$.

This result shows that the limiting experiment is a *locally asymptotically normal (LAN)* one in all three asymptotic scenarios; in the sequel, we will rather simply talk of *regimes*. While the central sequence is common to all regimes, the contiguity rate ν_n and information matrix Γ depend on the underlying regime[2]. As already mentioned, contiguity rates characterize how easy/hard it is to make inference on θ in each regime, hence are objects of interest. To comment on the dependence of contiguity rates on regimes, let us restrict to the case for which $r_n \equiv r$ (the radius r_n is indeed involved in the same way in each regime, in a fashion that, as it would be the case for a classical scale parameter, makes inference on θ easier if r_n converges to zero / harder if (r_n) diverges to infinity). In the classical regime (ii) where the noise magnitude is asymptotically fixed, the standard parametric contiguity rate $\nu_n = 1/\sqrt{n}$ is then obtained. As expected, the contiguity rate $\nu_n = \sigma_n/\sqrt{n}$ is faster in the favourable regime (i) where observations get closer and closer to the hypersphere. In view of the discussion in the introduction, it may come as a surprise, though, that the contiguity rate in the adverse regime (iii) is also

[1] Section S.n, Lemma S.$m.n$ or Equation (S.n) refer to the supplementary material.

[2] To keep the notation as light as possible, we will not write $\nu_{n(i)}$, $\nu_{n(ii)}$, $\nu_{n(iii)}$, nor $\Gamma_{(i)}$, $\Gamma_{(ii)}$, $\Gamma_{(iii)}$, for the contiguity rates and Fisher information matrices associated with the various regimes; we will rather make clear in each case what are the contiguity rates and information matrices we are referring to.

faster than the standard one in regime (ii). This may be related to the fact that, in the limit, regime (iii) actually puts half the probability mass (the probability mass inside the hypersphere) at θ, which is of course favourable (although the other half of the probability mass diverges to infinity). In any case, it is remarkable that, still in regime (iii), the contiguity rate depends on whether $d = 2$ or $d \geq 3$, with a more favourable rate obtained in the latter case: in this regime, it is thus more challenging to make inference on θ for $d = 2$ than in higher dimensions.

It is worth noticing that Theorem 2.1 holds under very mild assumptions. Throughout, the radius sequence (r_n) is completely arbitrary. As for the noise magnitude sequence (σ_n^2), it is also arbitrary in regimes (i)–(ii) (that is, there are no restrictions beyond those defining these regimes). In contrast, the result in regime (iii) imposes that $\exp((4+\varepsilon)\sigma_n^2) = o(n)$ for some $\varepsilon > 0$, a condition that, unless otherwise mentioned, we will always tacitly assume when considering this regime. Restricting again to the case $r_n \equiv r$ for the sake of simplicity, we have $\mathrm{Var}[X_{n1}] = c_n I_d$, with $c_n = (r^2/d)\exp(2\sigma_n^2)$, so that Theorem 2.1(iii) allows the variance-covariance matrix of X_{n1} (or more precisely, c_n) to diverge to infinity almost as fast as $\sqrt{n}$. In our results, the asymptotic normality of the central sequence $\Delta_n(\theta)$ is fully characterised in terms of the asymptotic behavior of σ_n^2 (see Proposition S.2.3). For example, if the variance-covariance matrix of X_{n1} diverges to infinity at least as fast as n (more precisely, if $\exp(2\sigma_n^2)$ is not $o(n)$), then the central sequence $\Delta_n(\theta)$ is not asymptotically normal, so that Theorem 2.1(iii) does *not* hold. However, the asymptotic normality of $\Delta_n(\theta)$ is guaranteed if the variance-covariance matrix of X_{n1} diverges to infinity at a rate between $\sqrt{n}$ and n. It remains an open question whether or not the asymptotic quadratic decomposition (2.2), hence also Theorem 2.1(iii), holds in this case.

Most importantly, the LAN result in Theorem 2.1 paves the way to defining optimal inference procedures on θ, which is one of our main objectives. We start with hypothesis testing.

3. Adaptively optimal hypothesis testing

In this section, we tackle the problem of testing the null hypothesis $\mathcal{H}_{n0} : \theta = \theta_0$ against the alternative hypothesis $\mathcal{H}_{n1} : \theta \neq \theta_0$, where θ_0 is a fixed d-vector. In Section 3.1, we construct the proposed test and study its asymptotic null behavior. Then, in Section 3.2, we derive its asymptotic behavior under contiguous alternatives and compare its performances to those of classical tests that ignore the hyperspherical a priori information.

3.1. The proposed test

The LAN property in Theorem 2.1 naturally provides optimal tests for the problem above. In particular, a direct consequence of this result is that, in the standard regime (ii) where $\sigma_n^2 \to \sigma^2 \in (0, \infty)$, the test, ϕ_n say, rejecting the null hypothesis $\mathcal{H}_{n0} : \theta = \theta_0$ at asymptotic level α when

$$
\begin{aligned}
Q_n &:= \Delta_n'(\theta_0)\Gamma^{-1}\Delta_n(\theta_0) \qquad (3.1)\\
&= \frac{d r_n^2}{n \exp(2\sigma_n^2)} \left((d-2)^2 + \frac{1}{\sigma_n^2} \right)^{-1} \sum_{i,j=1}^{n} \left(d + \frac{\log \|Z_{ni}(\theta_0)\| - \log r_n}{\sigma_n^2} \right) \\
&\qquad \times \left(d + \frac{\log \|Z_{nj}(\theta_0)\| - \log r_n}{\sigma_n^2} \right) \frac{Z_{ni}'(\theta_0) Z_{nj}(\theta_0)}{\|Z_{ni}(\theta_0)\|^2 \|Z_{nj}(\theta_0)\|^2}
\end{aligned}
$$

$$> \chi^2_{d,1-\alpha},$$

where Γ is the matrix in (2.3), is optimal in the Le Cam sense—more precisely, it is, for any $c > 0$, locally asymptotically maximin when testing $\mathcal{H}_{n0} : \theta = \theta_0$ against alternatives of the form $\mathcal{H}_{n1} : \|\theta - \theta_0\| \geq c$.

There is obviously no guarantee that this test meets the asymptotic level-α constraint in the non-standard regimes (i) and (iii) ("validity-robustness"). And if it does, there is also no guarantee that it remains Le Cam optimal in those non-standard regimes ("efficiency-robustness"). Quite remarkably, the test ϕ_n is actually both validity- and efficiency-robust in this sense, which is a direct corollary of the following result (see Section S.3 for a proof).

Theorem 3.1. *(a) In all regimes (i)–(iii), we have that, under* $\mathrm{P}^n_{\theta_0,r_n,\sigma_n^2}$,

$$Q_n = \Delta_n'(\theta_0)\Gamma^{-1}\Delta_n(\theta_0) + o_{\mathrm{P}}(1),$$

where Γ *is the information matrix associated with the corresponding underlying regime. (b) Under the null hypothesis* $\mathcal{H}_{n0} : \theta = \theta_0$*, the test statistic* Q_n *therefore remains asymptotically chi-square with* d *degrees of freedom both in regimes (i) and (iii), and, in these regimes, the test* ϕ_n *remains, for any* $c > 0$*, locally asymptotically maximin at level* α *when testing* $\mathcal{H}_{n0} : \theta = \theta_0$ *against* $\mathcal{H}_{n1} : \|\theta - \theta_0\| \geq c$.

This result is of course important since, in practice, the statistician will not know what is the underlying regime, so that adaptivity of ϕ_n in the underlying regime is a most desirable property. An important issue, however, is that ϕ_n involves the unspecified parameters r_n and σ_n^2, hence is an "oracle" test only. To obtain a feasible test, it is natural to replace these unspecified parameters with suitable estimators. Since $\log \|Z_{ni}(\theta_0)\| \sim \mathcal{N}(\log r_n, \sigma_n^2)$ under the null hypothesis, natural estimators are $\hat{r}_n(\theta_0)$ and $\hat{\sigma}_n^2(\theta_0)$, where we let

$$\hat{r}_n(\theta) := \exp\left(\frac{1}{n}\sum_{i=1}^n \log\|Z_{ni}(\theta)\|\right) = \sqrt[n]{\prod_{i=1}^n \|Z_{ni}(\theta)\|} \tag{3.2}$$

and

$$\hat{\sigma}_n^2(\theta) := \frac{1}{n}\sum_{i=1}^n \left(\log\|Z_{ni}(\theta)\| - \frac{1}{n}\sum_{i=1}^n \log\|Z_{ni}(\theta)\|\right)^2. \tag{3.3}$$

Irrespective of the underlying regime, we have that, under the null hypothesis $\mathcal{H}_{n0} : \theta = \theta_0$,

$$\frac{1}{\sigma_n}\log\left(\frac{\hat{r}_n(\theta_0)}{r_n}\right) = O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \frac{\hat{\sigma}_n^2(\theta_0)}{\sigma_n^2} - 1 = O_{\mathrm{P}}\left(\frac{1}{\sqrt{n}}\right);$$

see Lemma S.3.1. This suggests considering the test, $\hat{\phi}_n$ say, rejecting the null hypothesis at asymptotic level α when

$$\hat{Q}_n := \frac{d\hat{r}_n^2(\theta_0)}{n\exp(2\hat{\sigma}_n^2(\theta_0))}\left((d-2)^2 + \frac{1}{\hat{\sigma}_n^2(\theta_0)}\right)^{-1}\sum_{i,j=1}^n \left(d + \frac{\log\|Z_{ni}(\theta_0)\| - \log\hat{r}_n(\theta_0)}{\hat{\sigma}_n^2(\theta_0)}\right)$$
$$\times\left(d + \frac{\log\|Z_{nj}(\theta_0)\| - \log\hat{r}_n(\theta_0)}{\hat{\sigma}_n^2(\theta_0)}\right)\frac{Z_{ni}'(\theta_0)Z_{nj}(\theta_0)}{\|Z_{ni}(\theta_0)\|^2\|Z_{nj}(\theta_0)\|^2} > \chi^2_{d,1-\alpha}. \tag{3.4}$$

We have the following result (see Section S.3 for a proof).

Theorem 3.2. *In all regimes (i)–(iii),* $\hat{Q}_n = Q_n + o_{\mathrm{P}}(1)$ *under* $\mathrm{P}^n_{\theta_0, r_n, \sigma_n^2}$.

From Theorems 3.1 and 3.2, we conclude that the feasible test $\hat{\phi}_n$ is asymptotically equivalent, in any regime, to the oracle optimal test ϕ_n. This feasible test is thus not only validity-robust: it is also efficiency-robust in the sense that it is adaptively Le Cam optimal across all considered regimes. Of course, this asymptotic equivalence extends to sequences of contiguous alternatives, which will allow us to derive in the next section the corresponding asymptotic local powers of the feasible test in any regime.

3.2. Asymptotic non-null behavior

We consider local alternatives of the form $\mathcal{H}_{n1} : \theta_n = \theta_0 + \nu_n \tau$, where τ is a fixed non-zero d-vector and where ν_n is the contiguity rate of the underlying regime; see Theorem 2.1. The following result, that provides the asymptotic distribution of $\hat{Q}_n$ under such sequences of alternatives, of course trivially allows one to obtain the resulting local asymptotic powers.

Theorem 3.3. *In all regimes (i)–(iii), we have that, under* $\mathrm{P}^n_{\theta_0+\nu_n\tau, r_n, \sigma_n^2}$*, where* ν_n *is the contiguity rate associated with the underlying regime,* $\hat{Q}_n$ *is asymptotically non-central chi-square with* d *degrees of freedom and non-centrality parameter* $\|\tau\|^2/d$ *in regime (i),*

$$\frac{\exp(2\sigma^2)}{d}\left((d-2)^2 + \frac{1}{\sigma^2}\right)\|\tau\|^2$$

in regime (ii), $\|\tau\|^2/d$ *in regime (iii) with* $d = 2$*, and* $(d-2)^2\|\tau\|^2/d$ *in regime (iii) with* $d \geq 3$.

For the sake of comparison, we now study the corresponding asymptotic non-null behaviors of two classical tests that ignore the hyperspherical a priori information, namely the Hotelling test and the spatial sign test. Recall that the Hotelling test rejects the null hypothesis $\mathcal{H}_{n0} : \theta = \theta_0$ at asymptotic level α when

$$T_n^2 := n(\bar{X}_n - \theta_0)' S_n^{-1} (\bar{X}_n - \theta_0) > \chi^2_{d,1-\alpha},$$

where $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_{ni}$ and $S_n := \frac{1}{n}\sum_{i=1}^n (X_{ni} - \bar{X}_n)(X_{ni} - \bar{X}_n)'$ are the sample mean and sample covariance matrix, respectively. As for the spatial sign test[3] (see, e.g., Chapter 6 in Oja (2010)), it rejects the null hypothesis at asymptotic level α when

$$W_n^2 := n\bar{U}_n'(\theta_0) M_n^{-1} \bar{U}_n(\theta_0) > \chi^2_{d,1-\alpha},$$

where $\bar{U}_n(\theta_0) := \frac{1}{n}\sum_{i=1}^n U_{ni}(\theta_0)$ involves the "spatial signs" $U_{ni}(\theta_0) = Z_{ni}(\theta_0)/\|Z_{ni}(\theta_0)\|, i = 1, \ldots, n$ of the original observations with respect to θ_0, and where we let $M_n := \frac{1}{n}\sum_{i=1}^n U_{ni}(\theta_0)U_{ni}'(\theta_0)$. The following results then describe the null and non-null asymptotic behaviors of these tests in all regimes.

[3] For the sake of simplicity, we will simply refer to this test as the "sign test" in the sequel.

Theorem 3.4. *(a) In all regimes (i)–(iii), we have that, under* $\mathrm{P}^n_{\theta_0,r_n,\sigma_n^2}$*, the test statistic* T_n^2 *is asymptotically chi-square with* d *degrees of freedom; (b) in all regimes (i)–(iii), we have that, under* $\mathrm{P}^n_{\theta_0+\nu_n\tau,r_n,\sigma_n^2}$*, where* ν_n *is the contiguity rate associated with the underlying regime,* T_n^2 *is asymptotically chi-square with d degrees of freedom in regimes (i) and (iii), and non-central chi-square with d degrees of freedom and non-centrality parameter* $d\|\tau\|^2/\exp(2\sigma^2)$ *in regime (ii).*

Theorem 3.5. *(a) In all regimes (i)–(iii), we have that, under* $\mathrm{P}^n_{\theta_0,r_n,\sigma_n^2}$*, the test statistic* W_n^2 *is asymptotically chi-square with* d *degrees of freedom; (b) in all regimes (i)–(iii), we have that, under* $\mathrm{P}^n_{\theta_0+\nu_n\tau,r_n,\sigma_n^2}$*, where* ν_n *is the contiguity rate associated with the underlying regime,* W_n^2 *is asymptotically chi-square with d degrees of freedom in regimes (i) and (iii), and non-central chi-square with d degrees of freedom and non-centrality parameter* $(d-1)^2\exp(\sigma^2)\|\tau\|^2/d$ *in regime (ii).*

This shows that both the Hotelling test and the sign test are blind to contiguous alternatives in regimes (i) and (iii). Thus, in these regimes, taking into account the hyperspherical a priori information allows the detection of less severe alternatives (in terms of rate) that classical tests cannot see. In the standard regime (ii), on the contrary, these classical tests are rate-consistent, and their performances can then be compared to those of the proposed test through asymptotic relative efficiencies (AREs). As usual, AREs are obtained as ratios of the corresponding non-centrality parameters, so that, in regime (ii), the AREs of the feasible optimal test $\hat{\phi}_n$ with respect to the Hotelling test and with respect to the sign test are given by

$$\frac{\exp(4\sigma^2)}{d^2}\left((d-2)^2+\frac{1}{\sigma^2}\right) \quad \text{and} \quad \frac{\exp(\sigma^2)}{(d-1)^2}\left((d-2)^2+\frac{1}{\sigma^2}\right), \tag{3.5}$$

respectively. In Figure 1, we plot, for various dimensions d, the AREs in (3.5) as functions of the limiting noise magnitude σ^2. Note that these AREs are compatible with the AREs in regimes (i) and (iii), that may be considered infinite (all the curves in the right panel eventually diverge to infinity as σ^2 does). In any dimension d, the minimum of the AREs in (3.5) is strictly larger than one; the proposed test therefore always strictly improves over the performance of its classical competitors. For $d=2$, this minimal ARE is given by $e\approx 2.718$ both for the Hotelling test and for the sign test (the minimum is achieved at $\sigma^2=1/4$ for the Hotelling test and at $\sigma^2=1$ for the sign test).

4. Adaptively optimal estimation

We turn to estimation of θ. In the LAN framework of Theorem 2.1, an estimator $\hat{\theta}_n$ will be considered asymptotically optimal (see, e.g., Chapters 7 and 8 in van der Vaart (1998)) in a given regime if, for any θ, we have that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,

$$\nu_n^{-1}(\hat{\theta}_n-\theta)=\Gamma^{-1}\Delta_n(\theta)+o_{\mathrm{P}}(1), \tag{4.1}$$

where ν_n and Γ are respectively the contiguity rate and information matrix associated with this regime. From Theorem 2.1, $\nu_n^{-1}(\hat{\theta}_n-\theta)$ is then asymptotically normal with mean zero and covariance matrix Γ^{-1} (intuitively, asymptotic efficiency is thus related with the fact consistency is achieved at the optimal rate and that the variance-covariance matrix of the corresponding asymptotic distribution is equal to the inverse of the information matrix).

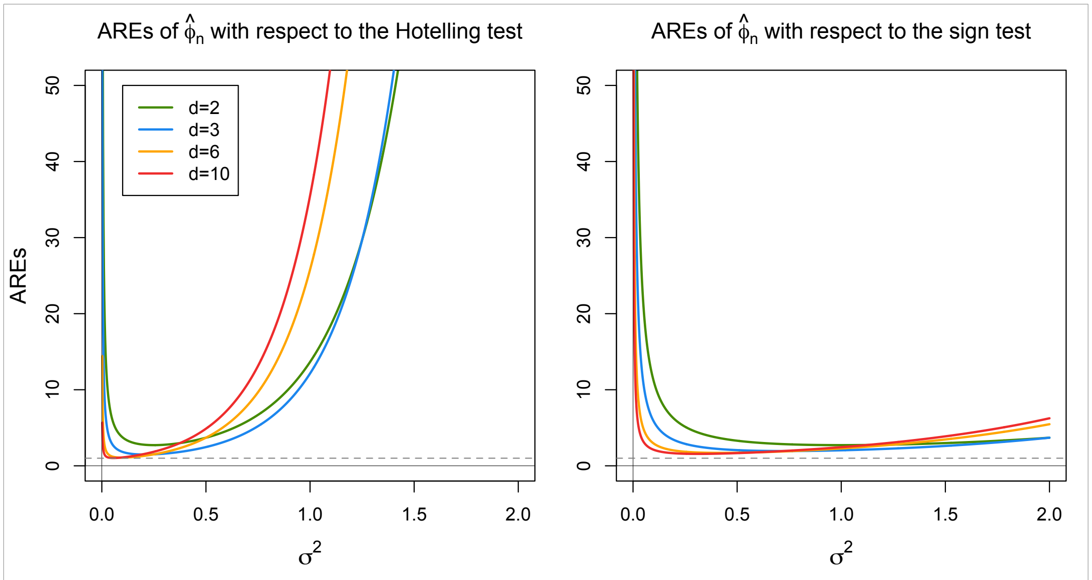

Figure 1. AREs, for dimension $d = 2, 3, 6, 10$, of $\hat{\phi}_n$ with respect to the Hotelling test (left) and with respect to the sign test (right), as a function of the limiting noise magnitude σ^2 in regime (ii) (the corresponding AREs may be considered infinite in regimes (i) and (iii)). In each panel, the dashed horizontal line shows the value one.

In the sequel, we will say that $\tilde{\theta}_n$, a preliminary estimator of θ, is *rate-consistent* in a given regime if and only if, for any θ, we have that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,

$$\nu_n^{-1}(\tilde{\theta}_n - \theta) = O_{\mathrm{P}}(1),$$

still with the contiguity rate ν_n associated with this regime. Based on such a preliminary estimator, one can consider the one-step estimator

$$\hat{\theta}_n = \tilde{\theta}_n + \hat{\nu}_n \hat{\Gamma}_n^{-1} \hat{\Delta}_n(\tilde{\theta}_n), \tag{4.2}$$

where, writing $\hat{r}_n := \hat{r}_n(\tilde{\theta}_n)$ and $\hat{\sigma}_n^2 := \hat{\sigma}_n^2(\tilde{\theta}_n)$ (see (3.2)–(3.3)), we let

$$\hat{\nu}_n := \frac{\hat{r}_n}{\sqrt{n}}, \qquad \hat{\Gamma}_n = \frac{\exp(2\hat{\sigma}_n^2)}{d}\left((d-2)^2 + \frac{1}{\hat{\sigma}_n^2}\right) I_d, \tag{4.3}$$

and

$$\hat{\Delta}_n(\theta) := \hat{\nu}_n \sum_{i=1}^{n} \left(d + \frac{\log\|Z_{ni}(\theta)\| - \log \hat{r}_n}{\hat{\sigma}_n^2}\right) \frac{Z_{ni}(\theta)}{\|Z_{ni}(\theta)\|^2}.$$

Again, the statistician will not know what is the underlying regime, which is the reason why, irrespective of the underlying regime, the quantities $\hat{\nu}_n$ and $\hat{\Gamma}_n$ in (4.3) to be used in the one-step estimator (4.2) are throughout associated with a *fixed* regime, namely the classical regime (ii)—this is in line with what was done in (3.1) and (3.4) for hypothesis testing, where the test statistics are quadratic forms that are throughout based on the Fisher information matrix associated with regime (ii).

The following result is the key to study the asymptotic behavior of the one-step estimator above (see Section S.4 for a proof).

Proposition 4.1. *In any regime under which $\tilde{\theta}_n$ is rate-consistent, we have for any θ that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,*

$$(a)\quad \nu_n \hat{\nu}_n^{-1} \hat{\Delta}_n(\tilde{\theta}_n) - \nu_n \hat{\nu}_n^{-1} \hat{\Delta}_n(\theta) = -\nu_n^{-1} \Gamma(\tilde{\theta}_n - \theta) + o_{\mathrm{P}}(1),$$

$$(b)\quad \nu_n \hat{\nu}_n^{-1} \hat{\Delta}_n(\theta) - \Delta_n(\theta) = o_{\mathrm{P}}(1),$$

and

$$(c)\quad \nu_n^2 \hat{\nu}_n^{-2} \hat{\Gamma}_n = \Gamma + o_{\mathrm{P}}(1),$$

where ν_n and Γ are the contiguity rate and information matrix associated with the underlying regime.

It readily follows from this result that, in any regime under which $\tilde{\theta}_n$ is rate-consistent, we have for any θ that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,

$$\begin{aligned}
\nu_n^{-1}(\hat{\theta}_n - \theta) &= \nu_n^{-1}(\tilde{\theta}_n - \theta) + \nu_n^{-1} \hat{\nu}_n \hat{\Gamma}_n^{-1} \hat{\Delta}_n(\tilde{\theta}_n) \\
&= \nu_n^{-1}(\tilde{\theta}_n - \theta) + \left(\nu_n^{-2} \hat{\nu}_n^2 \hat{\Gamma}_n^{-1}\right)\left(\nu_n \hat{\nu}_n^{-1} \hat{\Delta}_n(\tilde{\theta}_n)\right) \\
&= \nu_n^{-1}(\tilde{\theta}_n - \theta) + \Gamma^{-1}\left(\Delta_n(\theta) - \nu_n^{-1} \Gamma(\tilde{\theta}_n - \theta)\right) + o_{\mathrm{P}}(1) \\
&= \Gamma^{-1} \Delta_n(\theta) + o_{\mathrm{P}}(1).
\end{aligned}$$

This shows that the asymptotic behavior in probability of the one-step estimator in (4.2) does not depend on the particular preliminary estimator that is used, and that this behavior is the optimal one in (4.1). The consistency rate of the one-step estimator is the standard one in the classical regime (ii), but it coincides with the faster rates that were obtained in regimes (i) and (iii) for hypothesis testing.

Jointly with the asymptotic normality of $\Delta_n(\theta)$ in Theorem 2.1, we actually proved the following result.

Theorem 4.1. *In any regime under which $\tilde{\theta}_n$ is rate-consistent, we have for any θ that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,*

$$\nu_n^{-1}(\hat{\theta}_n - \theta) = \Gamma^{-1} \Delta_n(\theta) + o_{\mathrm{P}}(1) \to_{\mathcal{D}} \mathcal{N}(0, \Gamma^{-1}),$$

where ν_n and Γ are the contiguity rate and information matrix associated with the underlying regime.

While this result takes a form that may seem classical in Le Cam's one-step estimation, we emphasize that the multi-regime structure it covers, as well as the implicit requirement that the preliminary estimator exhibits different consistency rates across regimes, makes this result highly non-standard. It should also be noted that, remarkably, this result does not require that the preliminary estimator is "locally asymptotically discrete", which is a classical assumption in the framework of one-step estimation under local asymptotic normality; see, e.g., Kreiss (1987), Hallin, Oja and Paindaveine (2006), Ilmonen and Paindaveine (2011), or Hallin, Paindaveine and Verdebout (2014). Note also that Theorem 4.1 and

Proposition 4.1(c) entail that, in any regime under which $\tilde{\theta}_n$ is rate-consistent, we have for any θ that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,

$$\begin{aligned}\hat{\nu}_n^{-2}(\hat{\theta}_n-\theta)'\hat{\Gamma}_n(\hat{\theta}_n-\theta) &= (\nu_n^{-1}(\hat{\theta}_n-\theta))'(\nu_n^2\hat{\nu}_n^{-2}\hat{\Gamma}_n)(\nu_n^{-1}(\hat{\theta}_n-\theta)) \\ &= (\nu_n^{-1}(\hat{\theta}_n-\theta))'\Gamma(\nu_n^{-1}(\hat{\theta}_n-\theta))+o_{\mathrm{P}}(1) \\ &\to_{\mathcal{D}} \chi^2_d.\end{aligned}$$

A direct corollary is that, in any regime under which $\tilde{\theta}_n$ is rate-consistent, the hyper-ellipsoid

$$\left\{\theta\in\mathbb{R}^d : (\theta-\hat{\theta}_n)'\hat{\Gamma}_n(\theta-\hat{\theta}_n)\leq \hat{\nu}_n^2\chi^2_{d,1-\alpha}\right\}$$

is an asymptotic confidence zone for θ at confidence level $1-\alpha$. Thus, the construction above does not only provide a solution to point estimation but also to confidence zone estimation.

Theorem 4.1 shows that adaptively optimal estimation can be achieved provided that a preliminary estimator $\tilde{\theta}_n$ that is rate-consistent in all regimes is available. A natural choice is the sample mean $\bar{X}_n=\frac{1}{n}\sum_{i=1}^n X_{ni}$. However, a relatively straightforward application of the Lindeberg CLT shows that, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,

$$\frac{\sqrt{n}}{r_n\exp(\sigma_n^2)}(\bar{X}_n-\theta)\to_{\mathcal{D}}\mathcal{N}_d\Big(0,\frac{1}{d}I_d\Big);$$

see (S.3.3). Consequently, this preliminary estimator is rate-consistent in regime (ii), but not in regimes (i) and (iii). This discards this choice of $\tilde{\theta}$: again, since we do not know what is the underlying regime, we should indeed identify a preliminary estimator that is rate-consistent in as many regimes as possible.

We now construct a preliminary estimator that is rate-consistent in regimes (i)–(ii), hence can be used to perform adaptively optimal estimation in these regimes. For the sake of simplicity, write below $\varepsilon_{n1}=\log\|Z_{n1}(\theta)\|-\log r_n$ and $U_{n1}=U_{n1}(\theta)$, where θ is the true value of the location parameter. For any d-vector ϑ, we then have, under $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$,

$$\|Z_{n1}(\vartheta)\|^2=\|r_ne^{\varepsilon_{n1}}U_{n1}+(\theta-\vartheta)\|^2=r_n^2e^{2\varepsilon_{n1}}+\|\theta-\vartheta\|^2+2r_ne^{\varepsilon_{n1}}(\theta-\vartheta)'U_{n1},$$

hence also

$$\begin{aligned}\mathrm{Var}[\|Z_{n1}(\vartheta)\|^2] &= \mathrm{E}[\{\|Z_{n1}(\vartheta)\|^2-\mathrm{E}[\|Z_{n1}(\vartheta)\|^2]\}^2] \\ &= r_n^4\mathrm{Var}[e^{2\varepsilon_{n1}}]+\frac{4r_n^2}{d}\mathrm{E}[e^{2\varepsilon_{n1}}]\|\theta-\vartheta\|^2.\end{aligned}$$

Consequently,

$$\theta=\arg\min_{\vartheta\in\mathbb{R}^d}\mathrm{Var}[\|Z_{n1}(\vartheta)\|^2],$$

which makes it natural to consider the estimator

$$\tilde{\theta}_n:=\arg\min_{\vartheta\in\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^n\Big(\|Z_{ni}(\vartheta)\|^2-\frac{1}{n}\sum_{i=1}^n\|Z_{ni}(\vartheta)\|^2\Big)^2. \tag{4.4}$$

We have the following result.

Proposition 4.2. *(a) The estimator in (4.4) is uniquely defined (almost surely under any* $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$*) and admits the closed-form expression*

$$\tilde{\theta}_n = \frac{1}{2} S_n^{-1} T_n, \tag{4.5}$$

where S_n *still denotes the sample covariance matrix of* X_{ni}*,* $i = 1, \ldots, n$*, and where*

$$T_n := \frac{1}{n} \sum_{i=1}^{n} (d_{ni} - \bar{d}_n)(X_{ni} - \bar{X}_n)$$

is the sample covariance vector between $d_{ni} := \|X_{ni}\|^2$*,* $i = 1, \ldots, n$ *and* X_{ni}*,* $i = 1, \ldots, n$*. (b) This estimator is rate-consistent in regimes (i)–(ii).*

Jointly with Theorem 4.1, this result settles the problem of optimal estimation in regimes (i)–(ii). A close inspection of the proof of Proposition 4.2(b), however, reveals that the preliminary estimator in (4.4) is not rate-consistent in regime (iii). Existence of a preliminary estimator that would be rate-consistent in all three regimes remains an open question. Yet we conjecture that the estimator

$$\tilde{\theta}_n := \arg\min_{\vartheta \in \mathbb{R}^d} \hat{\sigma}_n^2(\vartheta)$$

(see (3.3)) is rate-consistent in all regimes (i)–(iii). Unfortunately, the lack of a closed-form expression makes the study of the asymptotic behavior of this estimator challenging in the non-standard regimes (i) and (iii). Even if this estimator is indeed rate-consistent in all regimes, the preliminary estimator in (4.4) remains the best solution if one restricts to the important regimes (i)–(ii), in view of its closed-form expression in Proposition 4.2(a).

5. Monte Carlo simulations

In this section, we perform several Monte Carlo exercises in order to investigate the finite-sample relevance of our asymptotic results. We first consider hypothesis testing (Section 5.1), then turn to point estimation (Section 5.2).

5.1. Hypothesis testing

Without any loss of generality, we focused throughout on the null hypothesis associated with $\theta_0 = 0$ and on alternatives for which τ is a multiple of e_1, the first vector of the canonical basis of $\mathbb{R}^d$. We started with Monte Carlo exercises in regime (i), where we took $r_n = 1$ and $\sigma_n^2 = 1/n^2$. We generated several collections of $M = 5000$ mutually independent d-variate samples from $\mathrm{P}^n_{\theta_0+\nu_n\tau,r_n,\sigma_n^2}$, for any combination of $n = \{100, 1000\}$, $d \in \{2, 6, 10\}$, and $\tau \in \{c_d \ell e_1 : \ell = 0, 1, 2, \ldots, 20\}$. The value $\ell = 0$ corresponds to the null hypothesis, whereas $\ell = 1, \ldots, 20$ provide increasingly severe alternatives; here, the values of c_d, namely $c_2 = .3$, $c_6 = .6$ and $c_{10} = .9$, were chosen in such a way that the largest asymptotic powers reached are roughly the same for the various dimensions d (the exact values of c_d are of course unimportant; we provide them for reproducibility purposes only). In each random sample, we then performed four tests at asymptotic level 5% for the null hypothesis $\mathcal{H}_{n0} : \theta = \theta_0$, namely (a) the oracle test ϕ_n, (b) its feasible version $\hat{\phi}_n$, (c) the sign test, and (d) the Hotelling test.

The resulting rejection frequencies are plotted in Figure 2, where the corresponding asymptotic power curves (obtained from Theorems 3.3–3.5) are also provided. Under the null hypothesis ($\ell = 0$), the oracle test, the feasible test, and the sign test have rejection frequencies close to the significance level of 5% in all considered scenarios. In contrast, the Hotelling test is liberal for the small sample size in dimensions $d = 6$ and $d = 10$. For $n = 1000$, all four tests show null rejection frequencies that agree with the significance level of 5%, in line with our asymptotic results. Turning to power behavior, for $n = 1000$, empirical power curves perfectly agree with asymptotic power curves in the present regime (i), which supports our theoretical results and confirms that the sign and Hotelling tests are blind to contiguous alternatives in this regime.

We turn to regime (ii), where we took $r_n = 1$ and $\sigma_n^2 = 0.01$. We still generated collections of $M = 5000$ mutually independent d-variate samples from $\mathrm{P}^n_{\theta_0+\nu_n\tau,r_n,\sigma_n^2}$ as above, now for any combination of $n = \{100, 1000\}$, $d \in \{2, 6, 10\}$, and $\tau \in \{c_d \ell e_1 : \ell = 0, 1, \ldots, 20\}$, with $c_2 = .03$, $c_6 = .06$ and $c_{10} = .07$. The resulting empirical and asymptotic power curves of the same four tests are plotted in Figure 3. In terms of calibration, results are similar to regime (i): the oracle test, feasible test, and sign test show null rejection frequencies that agree with the significance level of 5% in all cases, whereas the Hotelling test over-rejects for the smallest sample size and $d = 6$ and $d = 10$. For $n = 1000$, this calibration problem is solved and all four tests show null rejection frequencies similar to 5%. Regarding power behavior, the agreement between empirical and asymptotic power curves of the proposed oracle and feasible tests is here good even at the small sample size $n = 100$, and both curves match perfectly for the larger sample size, which again confirms our asymptotic results. Even in this framework where competing tests show non-trivial powers against contiguous alternatives, their performances are poor compared to those of the proposed tests, particularly so for the sign test.

Finally, we consider regime (iii), with $r_n = 1$ and $\sigma_n^2 = (\log n)/5$ (which is compatible with the condition that $\exp((4 + \varepsilon)\sigma_n^2) = o(n)$ for some $\varepsilon > 0$). We again generated collections of $M = 5000$ mutually independent d-variate samples from $\mathrm{P}^n_{\theta_0+\nu_n\tau,r_n,\sigma_n^2}$ as above, now for any combination of $n = \{100, 1000, 10^6\}$, $d \in \{2, 6, 10\}$, and $\tau \in \{c_d \ell e_1 : \ell = 0, 1, \ldots, 20\}$, with $c_2 = 1.5$, $c_6 = .8$ and $c_{10} = .55$. The resulting empirical and asymptotic power curves of the same four tests are plotted in Figure 3. Regarding calibration in this regime, all four tests show null rejection frequencies close to 5%, except for the smallest sample size and $d = 10$, where the Hotelling, oracle, and feasible tests are very slightly liberal. This deviation is rapidly corrected as sample size gets larger, and all tests have null rejection frequencies very close to 5% for $n = 1000$ and $n = 10^6$, confirming again our asymptotic results. Turning to power behavior, once more, empirical and asymptotic power curves agree for the largest sample size, but clearly the asymptotics seems to kick in more slowly than in regimes (i)–(ii) (and in particular, the agreement is still not great for $d = 2$ at the largest sample size considered here). The results in particular confirm that the sign and Hotelling tests are asymptotically blind to contiguous alternatives in regime (iii). It should be noted that, not only in the present regime (iii) but also in both other regimes, the difference in empirical power curves between the oracle and feasible tests is very small for $n = 100$ and almost nil for larger sample sizes.

5.2. Point estimation

For point estimation, we considered three setups, namely (i) $\sigma_n^2 = 1/n^2$, (i)′ $\sigma_n^2 = 1/n$ and (ii) $\sigma_n^2 = .1$ (again, $r_n = 1$ was used in each case). As the notation suggests, the first two setups correspond to

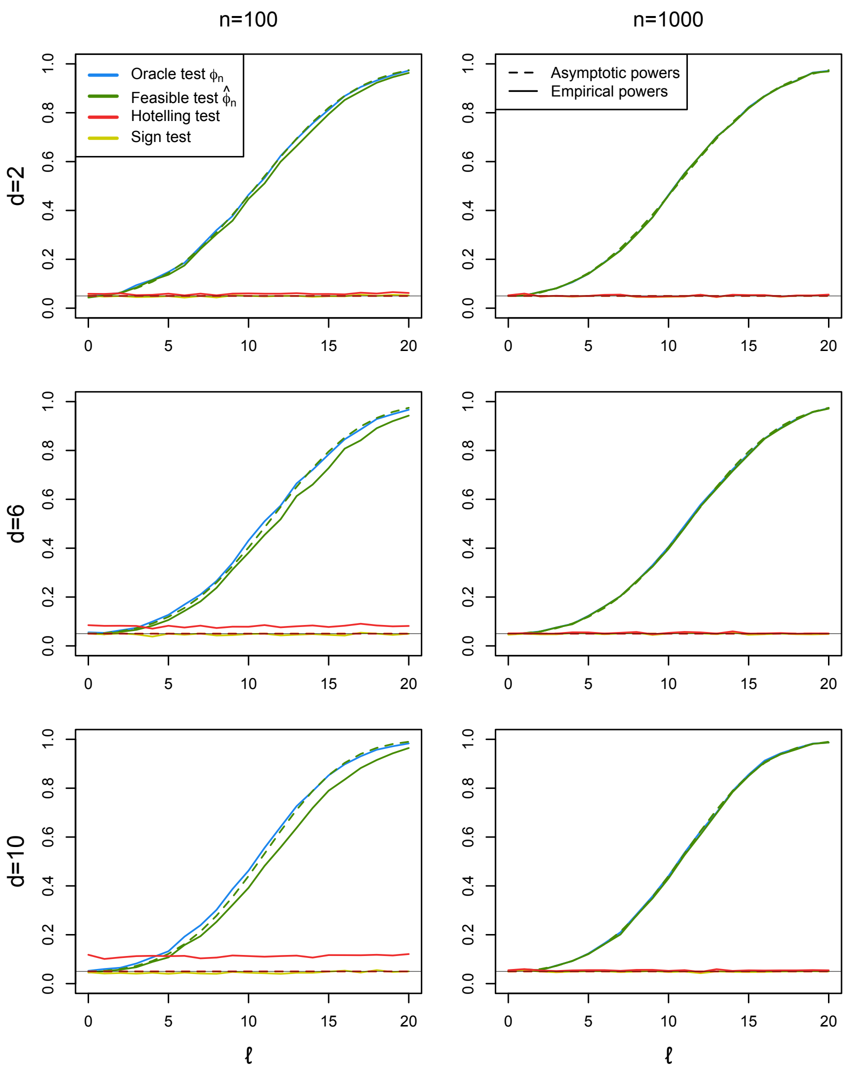

Figure 2. Empirical power curves (solid lines) of the oracle test ϕ_n, its feasible version $\hat{\phi}_n$, the sign test and the Hotelling test, obtained under null and non-null samples in regime (i): $\sigma_n^2 = 1/n^2$. Each empirical power is obtained from a collection of 5000 mutually independent samples; see Section 5.1 for details. The asymptotic power curves (dashed lines) are provided for the sake of comparison. The significance level of 5% (solid black line) is plotted for reference.

regime (i), whereas the third one is associated with regime (ii). In each setup, we fixed $\theta = 0$ (this is without loss of generality since all estimators we will consider are translation-equivariant) and we generated $M = 5000$ mutually independent d-variate random samples from $\mathrm{P}^n_{\theta,r_n,\sigma_n^2}$ for any combination of $n = \{100, 1000\}$ and $d \in \{2, 6, 10\}$. In each of the resulting samples, we then evaluated the following four estimators of θ: the sample mean $\bar{X}_n$, the one-step estimator $\hat{\theta}_n(\bar{X}_n)$ using the sample mean as a preliminary estimator, the preliminary estimator $\tilde{\theta}_n$ in (4.4), and the corresponding one-step estima-

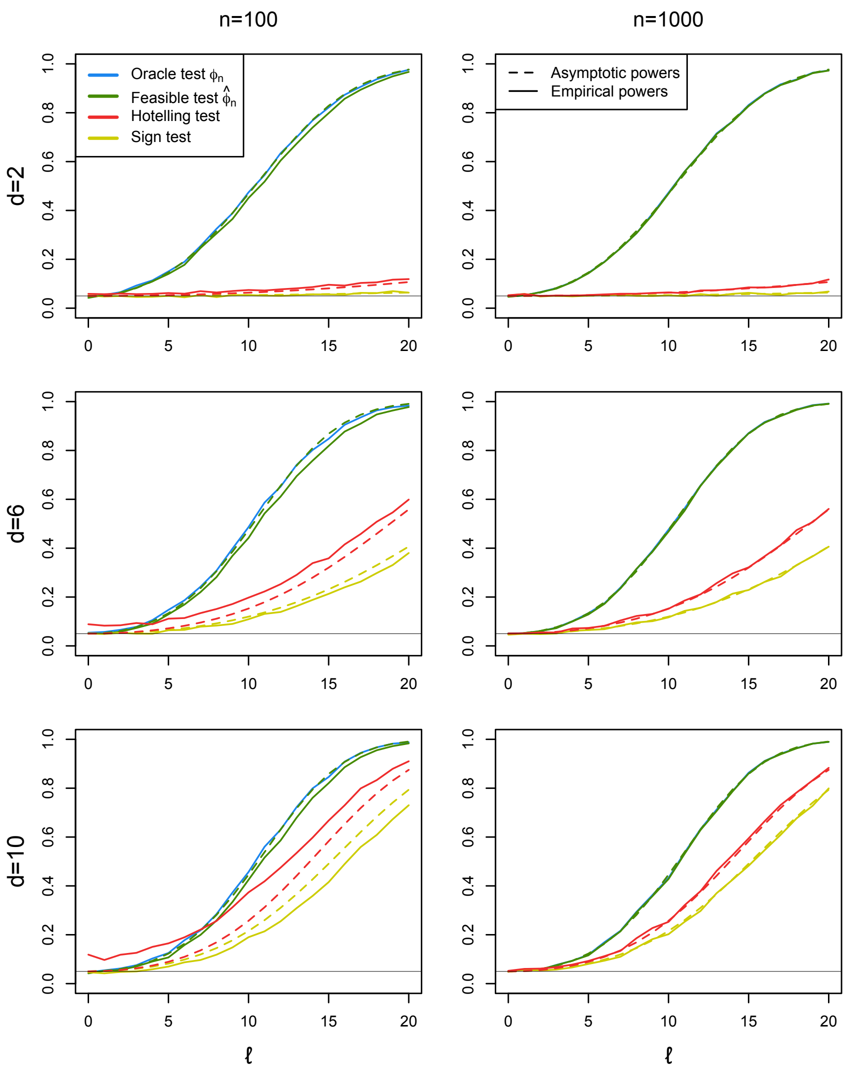

Figure 3. Empirical power curves (solid lines) of the oracle test ϕ_n, its feasible version $\hat{\phi}_n$, the sign test and the Hotelling test, obtained under null and non-null samples in regime (ii): $\sigma_n^2 = 0.01$. Each empirical power is still obtained from a collection of 5000 mutually independent samples; see Section 5.1 for details. The asymptotic power curves (dashed lines) are provided for the sake of comparison. The significance level of 5% (solid black line) is plotted for reference.

tor $\hat{\theta}_n$ ($\tilde{\theta}_n$). For each estimator, $\check{\theta}_n$ say, Figure 5 provides the boxplots of the corresponding normalized squared estimation errors

$$\|\nu_n^{-1}(\check{\theta}_n^{(m)} - \theta)\|^2, \qquad m = 1, \ldots, M, \tag{5.1}$$

where $\check{\theta}_n^{(m)}$ denotes the estimator $\check{\theta}_n$ evaluated in the mth generated random sample and where normalization is performed with respect to the contiguity rate ν_n associated with the underlying regime.

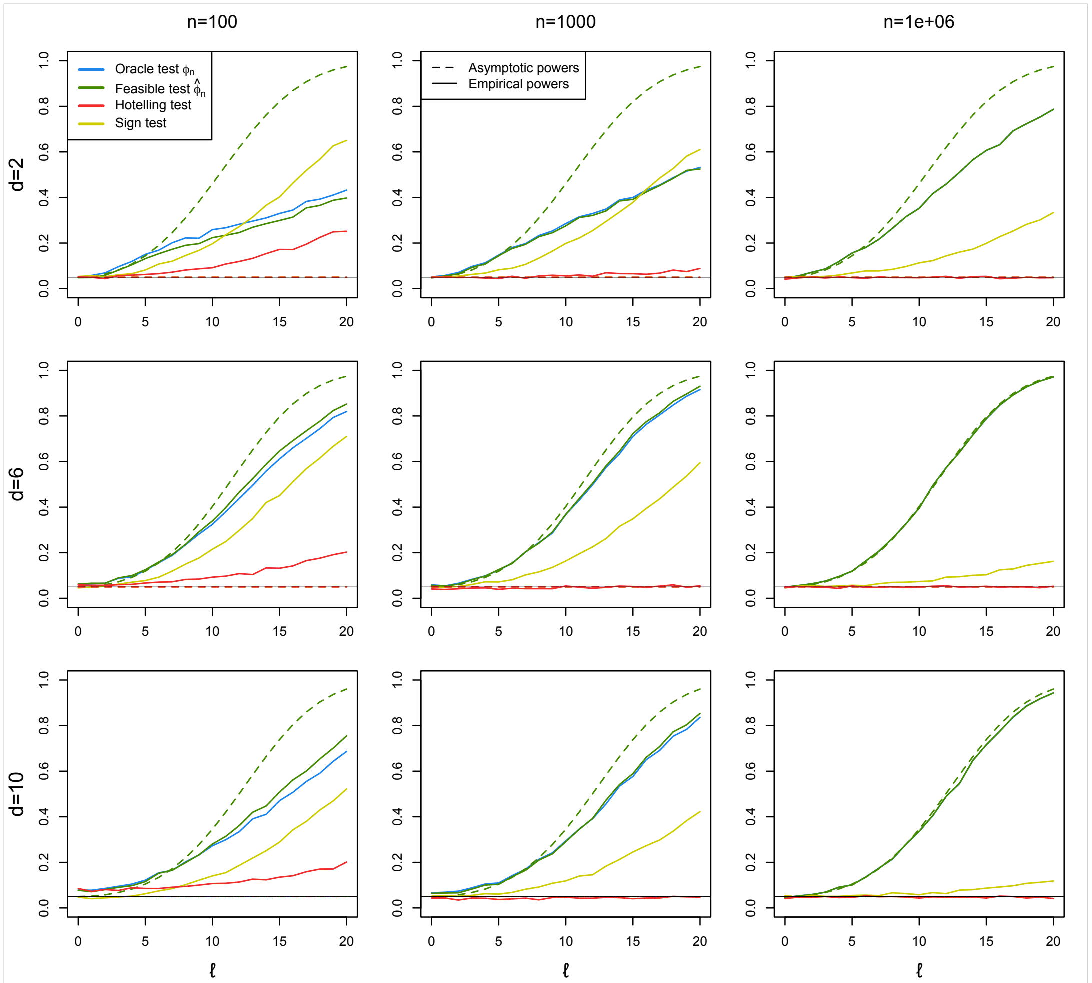

Figure 4. Empirical power curves (solid lines) of the oracle test ϕ_n, its feasible version $\hat{\phi}_n$, the sign test and the Hotelling test, obtained under null and non-null samples in regime (iii): $\sigma_n^2 = (\log n)/5$. Each empirical power is still obtained from a collection of 5000 mutually independent samples; see Section 5.1 for details. The asymptotic power curves (dashed lines) are provided for the sake of comparison. The significance level of 5% (solid black line) is plotted for reference.

The results are perfectly in line with our asymptotic results and further reveal interesting finite-sample behaviors:

- In setup (ii), both preliminary estimators are rate-consistent, hence can be used in the proposed one-step construction. It is seen that the error distributions of the corresponding one-step estimators do not depend on the preliminary estimator that is used, as predicted by our asymptotic theory. The results also show that the improvement compared to the preliminary estimator may be huge; this is the case with respect to $\bar{X}_n$ for $d = 2$ and with respect to the preliminary estimator $\tilde{\theta}_n$ in (4.4) in all dimensions (but impressively so in higher dimensions where this estimator is actually very poor).

- In setups (i) and (i)′, the sample mean is not rate-consistent, so that our asymptotic results remain silent about the behavior of the corresponding one-step estimator. And indeed it is seen that this one-step estimator behaves extremely poorly in Setup (i), and in particular does not provide the same error distribution as the one-step estimator based on the (rate-consistent) preliminary estimator in (4.4). The situation is better in setup (i)′, but this is only due to the fact that, at the finite-sample sizes considered, setup (i)′ is not so far from the setup (ii) where $\bar{X}_n$ is a perfectly valid preliminary estimator. Now, it is easy to see from the proof of Proposition 4.2 that, in regime (i), the limit, as n diverges to infinity, of the variance-covariance matrix of ($\nu_n^{-1}\times$) the preliminary estimator in (4.4) coincides with the inverse of the Fisher information matrix in regime (i). Recalling Theorem 4.1, this explains that, at least for the larger sample size $n = 1000$, no improvement seems to result from the one-step construction. Interestingly, an improvement is still obtained for $n = 100$ and $d = 10$ in setup (i)′.

In all cases, it is seen that the magnitude of normalized squared estimation errors is the same at both sample sizes considered, which is compatible with the fact that the consistency rate is indeed the one associated with the contiguity rate ν_n of the underlying regime. As a conclusion, these empirical findings indeed fully support our asymptotic results and indicate that the one-step estimator based on the preliminary estimator in (4.4) is, both in regimes (i) and (ii), an excellent estimator in the considered distributional framework.

6. Wrap up and perspectives for future research

In this work, we considered noisy directional data, that is, data that deviate from their natural hyperspherical sample space due to some radial noise, and we tackled inference problems on the centre θ of the unknown, latent, hypersphere. We here briefly summarize what our theoretical results revealed in this context: first, consistency rates depend crucially on whether the noise magnitude (i) converges to zero, (ii) converges to a positive value, or (iii) diverges to infinity.[4] As expected, consistency rates are faster in regime (i) than in regime (ii); yet surprisingly at first, they are also faster in regime (iii) than in regime (ii). Second, we could define *adaptively* optimal inference procedures, that is, procedures that achieve Le Cam optimality in all regimes (i)–(iii). Third, efficiency bounds are the same in both setups where the radius of the underlying manifold is specified or unspecified, that is, the unspecification of this radius has no asymptotic cost in terms of efficiency. Fourth and last, the hyperspherical a priori information is precious to perform inference on θ: in regime (ii), classical procedures that ignore this a priori information are rate-consistent yet pay a positive price in terms of efficiency, whereas, in the less standard regimes (i) and (iii), they are not even rate-consistent. This fully answers the questions we had raised in the introduction of the paper.

Perspectives for future research are rich and diverse. While our asymptotic investigation is extensive in the considered framework where the dimension d was fixed, one could think of tackling the corresponding high-dimensional situation where the dimension $d = d_n$ diverges to infinity with n. Of

[4] While contiguity rates also depend on the radius r_n, this quantity does not provide further regimes (it was unclear a priori that only σ_n would discriminate between the various asymptotic regimes, and this is thus another point our asymptotic investigation reveals).

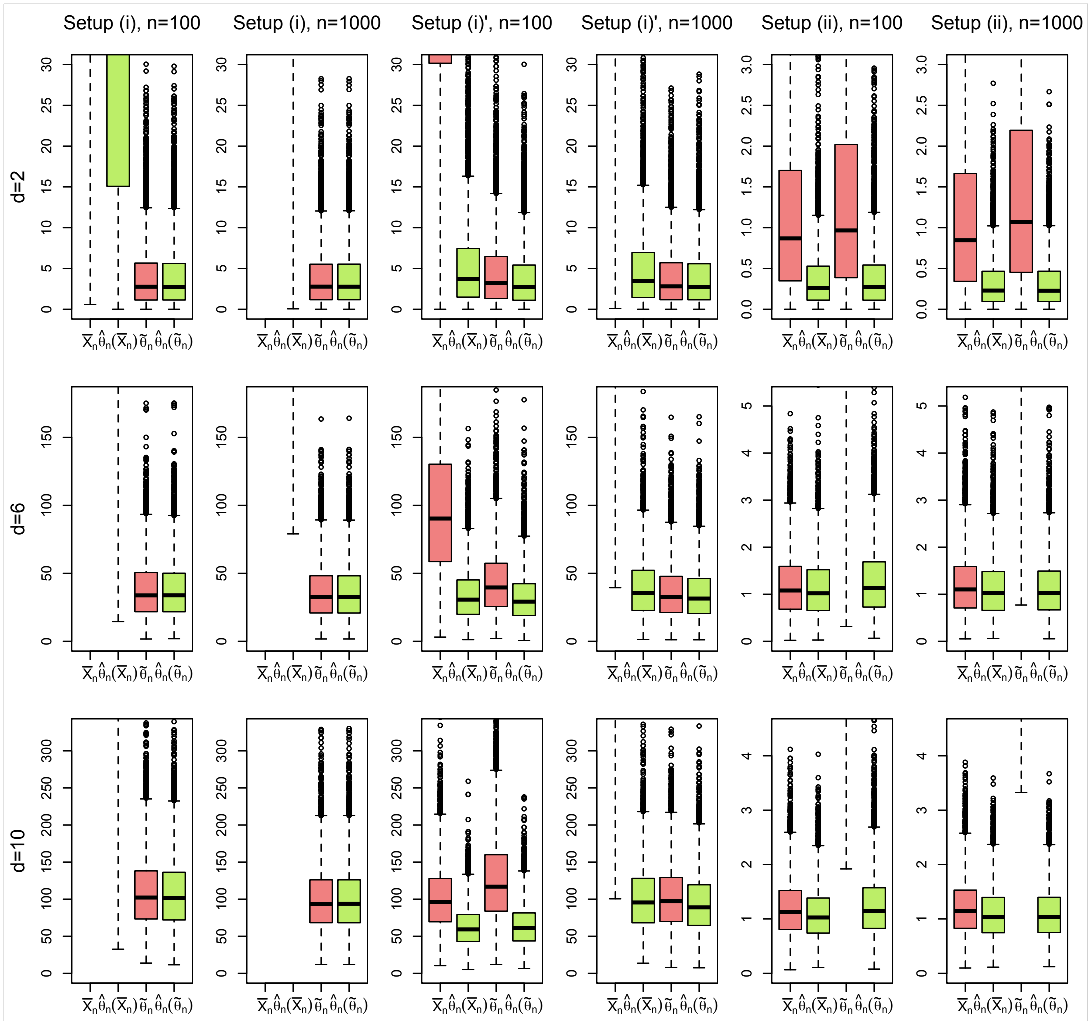

Figure 5. For any combination of a sample size $n = \{100, 1000\}$ and a dimension $d \in \{2, 6, 10\}$, this figure provides boxplots of the normalized squared estimation errors in (5.1) for four estimators obtained from 5000 mutually independent random samples generated with (i) $\sigma_n^2 = 1/n^2$, (i)′ $\sigma_n^2 = 1/n$ and (ii) $\sigma_n^2 = .1$. The estimators that are considered are the sample mean $\bar{X}_n$, the preliminary estimator $\tilde{\theta}_n$ in (4.4), and both corresponding one-step estimators $\hat{\theta}_n$; see Section 5.2 for details.

course, the dimension of the parameter of interest θ would then depend on n, which is in principle incompatible with Le Cam's asymptotic theory; in the spherically symmetric distributional framework adopted in this work, a reduction via invariance arguments, such as the one conducted in Paindaveine and Verdebout (2020a), might take care of this issue. Now, it is part of the folklore in high dimensions that radii (distances to the location centre) are eventually unimportant for inference on location and that, in this sense, directions from the centre become sufficient statistics. Since the only differences between regimes (i)–(iii) above are in terms of radii, this would suggest that, in sharp contrast with the low-dimensional setup, no gain would then be obtained from the hyperspherical a priori information in the high-dimensional setup. Whether this is the case or not is of course totally unclear and deserves a careful theoretical investigation, all the more so that the AREs of our optimal tests with respect to the sign test in (3.5) actually suggest otherwise (would the hyperspherical a priori information be useless in higher dimensions, then these AREs should typically converge to one as d diverges to infinity, which is not the case).

Back in the low-dimensional case, another possible avenue for future research is to broaden the distributional framework we restricted to in the present paper. In this first work on the topic, it was natural to consider a parametric framework to fix ideas. But of course, one might alternatively adopt a semiparametric model, in which the distribution of the radial noise would remain unspecified. To be more specific, the random variable ε_{ni} in (1.1) would then be assumed to admit a density of the form $z \mapsto \sigma_n^{-1} f(z/\sigma_n)$, where (in order to avoid moment assumptions:) f is a density with median zero and median absolute deviation one (this defines a slightly different noise magnitude quantity than in the parametric case considered in the present paper, but the difference is of course irrelevant). It would obviously be of high interest to identify which of our results extend to this semiparametric framework and, in contrast, which tightly depend on the Gaussian assumption we made in the present paper. Preliminary investigations suggest that the consistency rates obtained in regimes (i)–(ii) do not depend on f, whereas those in regime (iii) do. This semiparametric framework also raises the problem of defining *doubly-adaptive* inference procedures, that would combine semiparametric adaptivity (that is, adaptivity in the unknown f) with the adaptivity in the noise magnitude σ_n^2 we considered in this paper. Finally, rank-based procedures would also be natural to consider in this unspecified-f framework. Like the high-dimensional case, however, this semiparametric extension calls for entirely different techniques, hence is left for future research work.

Acknowledgments

The authors would like to thank the Editor-In-Chief Prof. Kengo Kato, the Associate Editor, and two anonymous referees for their insightful comments and suggestions on the original version of the paper. Davy Paindaveine is also affiliated at the Toulouse School of Economics, Université Toulouse 1 Capitole.

Funding

The authors are supported by two “Projet de Recherche” grants from the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique. Diego Bolón also acknowledges the support from the Spanish Ministerio de Ciencia, Innovación y Universidades (fellowship FPU20/03960), from the project PID2020-116587GB-I00 supported by the Agencia Estatal de Investigación (AEI) of the Ministerio de Ciencia, Innovación y Universidades, and from the Competitive Reference Groups 2021-2024 (ED431C 2021/24) from Xunta de Galicia.

Supplementary Material

Supplement to “Inference on location for noisy directional data: A Le Cam approach to quantify the value of the hyperspherical a priori information"
In this supplement, we prove all results of the present paper.

References

AGOSTINELLI, C. and GRECO, L. (2019). Weighted likelihood estimation of multivariate location and scatter. *Test* **28** 756–784.

BOLÓN, D., PAINDAVEINE, D. and VERDEBOUT, T. (2025). Supplement to “Inference on location for noisy directional data: A Le Cam approach to quantify the value of the hyperspherical a priori information”. Online supplement to the present submission.

CHAKRABORTY, A. and CHAUDHURI, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann. Statist.* **45** 771–799.

CHEN, L. S., PAUL, D., PRENTICE, R. L. and WANG, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *J. Amer. Statist. Assoc.* **106** 1345–1360.

CHENG, X. and XIE, Y. (2024). Kernel two-sample tests for manifold data. *Bernoulli* **30** 2572–2597.

DONOHO, D. L. and GRIMES, C. (2003). Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA* **100** 5591–5596.

DÜRRE, A. and PAINDAVEINE, D. (2022). Affine-equivariant inference for multivariate location under L_p loss functions. *Ann. Statist.* **50** 2616–2640.

FENG, L., ZOU, C. and WANG, Z. (2016). Multivariate-sign-based high-dimensional tests for the two-sample location problem. *J. Amer. Statist. Assoc.* **111** 721–735.

FENG, L., ZOU, C., WANG, Z. and ZHU, L. (2017). Composite T^2 test for high-dimensional data. *Statist. Sinica* **27** 1419–1436.

FRAHM, G., NORDHAUSEN, K. and OJA, H. (2020). M-estimation with incomplete and dependent multivariate data. *J. Multivariate Anal.* **176** 104569.

HALLIN, M., OJA, H. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *Ann. Statist.* **34** 2757–2789.

HALLIN, M. and PAINDAVEINE, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133.

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2014). Efficient R-estimation of principal and common principal components. *J. Amer. Statist. Assoc.* **109** 1071–1083.

HOTELLING, H. (1931). The generalization of Student's ratio. *Ann. Math. Statist.* **2** 360–378.

ILMONEN, P. and PAINDAVEINE, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39** 2448–2476.

KOCK, A. B. and PREINERSTORFER, D. (2019). Power in high-dimensional testing problems. *Econometrica* **87** 1055–1069.

KOCK, A. B. and PREINERSTORFER, D. (2023). Consistency of p-norm based tests in high dimensions: Characterization, monotonicity, domination. *Bernoulli* **29** 2544–2573.

KREISS, J. P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.

LAROCQUE, D., NEVALAINEN, J. and OJA, H. (2007). A weighted multivariate sign test for cluster-correlated data. *Biometrika* **94** 267–283.

LE CAM, L. M. and YANG, G. L. (2000). *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media.

LEY, C., SWAN, Y., THIAM, B. and VERDEBOUT, T. (2013). Optimal R-estimation of a spherical location. *Statist. Sinica* **23** 305–333.

LI, H., AUE, A., PAUL, D., PENG, J. and WANG, P. (2020). An adaptable generalization of Hotelling's T^2 test in high dimension. *Ann. Statist.* **48** 1815–1847.

MAGGIONI, M., MINSKER, S. and STRAWN, N. (2016). Multiscale dictionary learning: non-asymptotic bounds and robustness. *J. Mach. Learn. Res.* **17** 43–93.

OJA, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks*. Springer-Verlag, New York.

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.

PAINDAVEINE, D. and VERDEBOUT, T. (2020a). Detecting the direction of a signal on high-dimensional spheres: non-null and Le Cam optimality results. *Probab. Theory Related Fields* **176** 1165–1216.

PAINDAVEINE, D. and VERDEBOUT, T. (2020b). Inference for spherical location under high concentration. *Ann. Statist.* **48** 2982–2998.

RANDLES, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.* **95** 1263–1268.

SANDOVAL, J., UENISHI, K., IWAKIRI, M. and TANAKA, K. (2020). Robust sphere detection in unorganized 3D point clouds using an efficient Hough voting scheme based on sliding voxels. *IIEEJ Trans. Image Electron. Vis. Comput.* **8** 121–135.

SHAPIRO, A., XIE, Y. and ZHANG, R. (2021). Goodness-of-fit tests on manifolds. *IEEE Trans. Inform. Theory* **67** 2539–2553.

VAN DER GLAS, M., VOS, F. M., BOTHA, C. P. and VOSSEPOEL, A. M. (2002). Determination of position and radius of ball joints. In *Medical Imaging: Image Processing*.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge.

WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669.

WRIGHT, J. and MA, Y. (2022). *High-Dimensional Data Analysis with Low-Dimensional Models. Principles, Computation, and Applications*, 1st ed. Cambridge University Press, Cambridge.