Journal of Multivariate Analysis 178 (2020) 104603

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Kurtosis test of modality for rotationally symmetric distributions on hyperspheres

Byungwon Kim [a], Jörn Schulz [b,c], Sungkyu Jung [a,*]

[a] Department of Statistics, Seoul National University, Seoul 08826, South Korea
[b] Department of Mathematics and Physics, University of Stavanger, 4036 Stavanger, Norway
[c] Department of Research, Stavanger University Hospital, 4068 Stavanger, Norway

ARTICLE INFO

Article history:
Received 15 November 2019
Received in revised form 20 February 2020
Accepted 21 February 2020
Available online 28 February 2020

AMS 2010 subject classifications:
primary 62H11
62H25
secondary 62E10

Keywords:
Ball uniform distribution
Directional data
Exponential map
Fréchet mean
Hausdorff moment
Stieltjes moment

ABSTRACT

A test of modality of rotationally symmetric distributions on hyperspheres is proposed. The test is based on a modified multivariate kurtosis defined for directional data on $\mathbb{S}^d$. We first reveal a relationship between the multivariate kurtosis and the types of modality for Euclidean data. In particular, the kurtosis of a rotationally symmetric distribution with decreasing sectional density is greater than the kurtosis of the uniform distribution, while the kurtosis of that with increasing sectional density is less. For directional data, we show an asymptotic normality of the modified spherical kurtosis, based on which a large-sample test is proposed. The proposed test of modality is applied to the problem of preventing overfitting in non-geodesic dimension reduction of directional data. The proposed test is superior than existing options in terms of computation times, accuracy and preventing overfitting. This is highlighted by a simulation study and two real data examples.

1. Introduction

Directional statistics concern data on the hypersphere $\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^{d+1} : \mathbf{x}^\top \mathbf{x} = 1\}$, and have gained much attention recently. Data on hyperspheres or on spaces relevant to hyperspheres include dihedral angles of protein structures on $\mathbb{S}^1 \times \mathbb{S}^1$, directions of astrophysical events on $\mathbb{S}^2$, pre-shapes of landmark-based object shapes on $\mathbb{S}^d$, skeletally-represented objects on $\mathbb{S}^{d_1} \times (\mathbb{S}^2)^{d_2}$, and standardized multivariate data on $\mathbb{S}^{d-1}$. We refer to [26] for an overview of the field.

We focus on rotationally symmetric distributions on $\mathbb{S}^d$, and investigate the modality patterns. Modality refers to the shape of a distribution related to its modes. In doing so, we first contemplate on the types of modality for rotationally symmetric distributions on $\mathbb{R}^d$, the Euclidean multivariate case. A distribution on $\mathbb{R}^d$ with the density function f is rotationally symmetric about $\boldsymbol{\mu} \in \mathbb{R}^d$ if $f(\mathbf{x}) = f\{R(\mathbf{x} - \boldsymbol{\mu}) + \boldsymbol{\mu}\}$ for any orthogonal matrix R. Similarly, a distribution on the hypersphere $\mathbb{S}^d$ with the density f is rotationally symmetric about $\boldsymbol{\mu} \in \mathbb{S}^d$ if $f(\mathbf{x}) = f(R\mathbf{x})$ for any orthogonal matrix R satisfying $R\boldsymbol{\mu} = \boldsymbol{\mu}$. Inference problems for the location parameter $\boldsymbol{\mu}$ of rotationally symmetric directional distributions have gained much attention recently [22,29,30]. An optimal test for rotational symmetry is studied in [11].

We categorize the modality of rotationally symmetric distributions into three typical types: a unique mode at the mean, mode everywhere on the support, and a circular modal ridge. In Fig. 1, the toy data on the top row exhibit the first

* Corresponding author.
E-mail address: sungkyu@snu.ac.kr (S. Jung).

https://doi.org/10.1016/j.jmva.2020.104603
0047-259X/

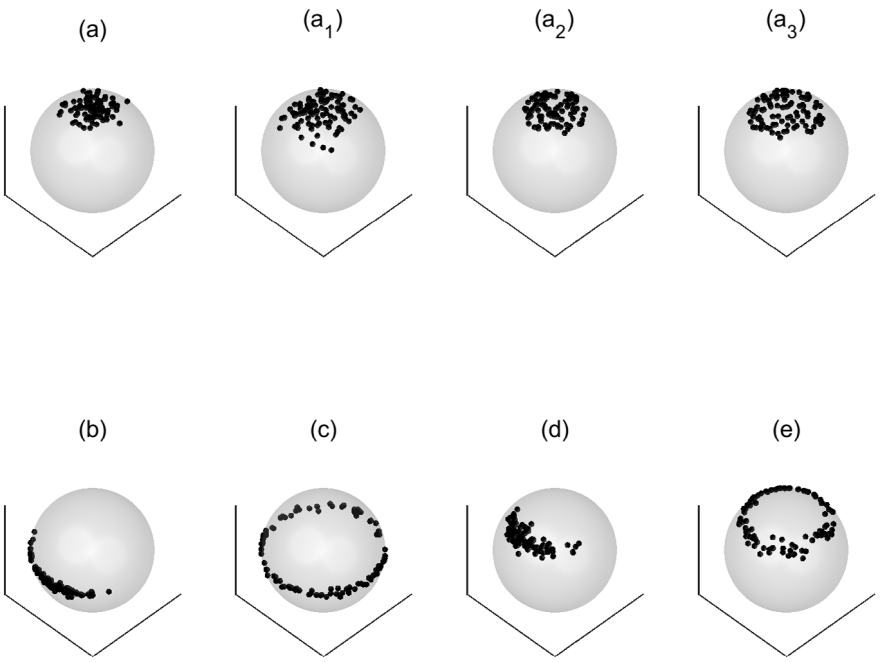

Fig. 1. Data examples on $\mathbb{S}^2$, sampled from von Mises–Fisher (a), a tangent-normal (a_1), a tangent uniform (a_2), a spherical ball uniform (a_3), small sphere distributions [17] (b) and (d), and Bingham–Mardia distributions [4] (c) and (e). See Sections 3.1 and 4.1.

two types of modality. On the other hand, samples in (c) and (e) show a clear circular ridge of modes. Samples in (b) and (d) are not rotationally symmetric and are discussed later. Our goal is to discriminate rotationally symmetric distributions with circular modal ridges from those with a unique mode.

Starting with the simpler Euclidean case, we propose to utilize and modify the multivariate kurtosis [25] in order to delineate different types of modalities. Write $\kappa(f)$ for the multivariate kurtosis of the distribution with density f. We show that if f is rotationally symmetric with a unique and single mode, then $\kappa(f) > \kappa(\mathcal{U}_d)$, where $\mathcal{U}_d$ is the uniform distribution on any rotationally symmetric and convex support in $\mathbb{R}^d$. In a class of alternative settings, we show that $\kappa(f) \leq \kappa(\mathcal{U}_d)$. Heuristically, the alternative settings include densities f with a ring-like ridge of modes. These facts, discussed in Section 2.1, justify converting hypotheses on the types of modality to hypotheses on $\kappa(f)$, provided that f is rotationally symmetric. A test procedure for modality is then built on the asymptotic normality of a sample kurtosis, modified for rotationally symmetric distributions.

For directional data, we use the tangent-space approximation to define and compute spherical kurtosis. The rotational symmetry of spherical distributions is preserved when the Exponential map is used for the approximation. See Section 3.2 for the definition of tangent space and Exponential map. It is argued that the spherical kurtosis is indicative of the types of modality for rotationally symmetric directional data. Using the fact that the infimum spherical kurtosis of unimodal distributions on $\mathbb{S}^d$ is again $\kappa(\mathcal{U}_d)$, we propose a test based on the asymptotic normality of a modified spherical kurtosis, and show that the test is consistent.

This work is motivated by a backward dimension reduction method for directional data on $\mathbb{S}^d$, proposed in [15,16], and analyzed and applied in various contexts [8,9,14,32,34,38]. To facilitate our discussion, Fig. 1 collects some typical data situations on $\mathbb{S}^2$. In [15], the dimension of data on $\mathbb{S}^2$ is reduced via a fitted curve, among the parametrized curves that are either great circles (geodesics, analogous to lines in $\mathbb{R}^2$) or small circles (a class of non-geodesic curves). In Fig. 1, cases (b) and (c) calls for geodesic dimension reduction, while a small circle fitting has an advantage for cases (d) and (e). In the cases (a)–(a_3) in the top row, fitting a small circle for such a data is an overfit. (An example of overfitted small circle is shown in the right panel of Fig. 4.) We apply the proposed test of modality in preventing the overfit. This new method performs superior and is computationally faster than existing options.

On technical sides, we believe that our results in Theorem 1 and Lemma 4 are nontrivial, and deserve separate attention. First, Theorem 1 relates the shape of a density with the kurtosis. While there is some literature on whether univariate kurtosis is indicative of the shape of a density, the literature on the meaning of multivariate kurtosis is relatively thin; see [18,21,37] for instance. Our work closes this gap. In particular, our proof utilizing the properties of Hausdorff and Stieltjes moment sequences suggests that the kurtosis of a distribution should be compared with that of the uniform distribution. Second, the result in Lemma 4 formalizes tangent-space approximated inference procedures, which have been used quite informally in the field of directional statistics.

The rest of paper is organized as follows. In Section 2, we formally connect Mardia's multivariate kurtosis with modality types, and propose a kurtosis test of modality for Euclidean rotationally symmetric multivariate data. The Euclidean notion of multivariate kurtosis and related arguments are extended to directional data in Section 3. In Section 4, we demonstrate that the proposed test, when applied to the overfitting problem of spherical dimension reduction, performs well. This is further highlighted in a dimension reduction for 3D models of human hippocampi and in an analysis of real astronomical data, in Section 5. Remarks on multivariate kurtosis and related discussions are contained in Section 6. Nearly all proofs are contained in Section 7.

We write $\mathbf{X} \sim f$ when the density of the random vector $\mathbf{X} = (X_1, \ldots, X_d)^\top$ is f, and $\|\mathbf{x}\|$ for the usual 2-norm of a vector $\mathbf{x}$. We assume that the distribution is absolutely continuous.

2. Method for Euclidean multivariate data

2.1. Multivariate kurtosis as a modality indicator

The kurtosis of a random variable $X \sim f$ is defined as the fourth standard moment $\kappa(f) = \mathrm{E}\{(X-\mu)^4/\sigma^4\}$, where $\mu = \mathrm{E}(X)$, $\sigma^2 = \mathrm{Var}(X)$. Generalizing the univariate notion of kurtosis to multivariate distributions, Mardia [25] defined the multivariate kurtosis $\kappa(f)$, for $\mathbf{X} \sim f_{\mathbf{X}}$,

$$\kappa(f_{\mathbf{X}}) = \mathrm{E}\{(\mathbf{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\}^2, \tag{1}$$

where $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X})$ and $\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X})$. It is straightforward to check that the multivariate kurtosis is invariant under nonsingular affine transformations of the random vector $\mathbf{X}$. Thus, for the standardized random vector $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu}) \sim f_{\mathbf{Z}}$, we have $\kappa(f_{\mathbf{X}}) = \kappa(f_{\mathbf{Z}})$ for any $f_{\mathbf{X}}$, and $\kappa(f_{\mathbf{Z}}) = \mathrm{E}(\mathbf{Z}^\top\mathbf{Z})^2 = \mathrm{E}\|\mathbf{Z}\|^4$. (We write $\mathrm{E}\|\mathbf{Z}\|^p$ for $\mathrm{E}(\|\mathbf{Z}\|^p)$.) In words, the multivariate kurtosis is the fourth moment of the distance of the standardized data points from the mean. The sample kurtosis will be discussed in Section 2.2.

In the following we argue that the multivariate kurtosis is indicative of the modality of the rotationally symmetric distribution. The mode of a distribution with the density function $f_{\mathbf{X}}$ is the location $\mathbf{x}_0 \in \mathbb{R}^d$ such that $f_{\mathbf{X}}(\mathbf{x}_0) \geq f_{\mathbf{X}}(\mathbf{x})$, for any $\mathbf{x} \in \mathbb{R}^d$. We list a few properties of rotationally symmetric distributions.

Lemma 1. *Let $\mathbf{X} \sim f_{\mathbf{X}}$ be rotationally symmetric about the origin in $\mathbb{R}^d$. Then*

(i) $\mathrm{E}(\mathbf{X}) = \mathbf{0}$ *and* $\mathrm{Cov}(\mathbf{X}) = \sigma^2\mathbf{I}_d$, *for some* $\sigma^2 > 0$.
(ii) *If $\mathbf{x}_0$ is a mode of $f_{\mathbf{X}}$, then for any orthogonal matrix R, $R\mathbf{x}_0$ is also a mode of $f_{\mathbf{X}}$.*
(iii) *If $\mathbf{x}_0 \in \mathbb{R}^d$ is the unique mode of $f_{\mathbf{X}}$, then $\mathbf{x}_0 = \mathbf{0}$.*

The proof of Lemma 1 is omitted. It is clear from Lemma 1(ii) and (iii) that there are no rotationally symmetric distributions having two or more (countable) number of modes. There are either one or infinitely many modes. To investigate the relation of the kurtosis to the modality of rotationally symmetric distributions, we take three typical modalities:

I. A unique mode at the center of the distribution. For example, the standard multivariate normal distribution.
II. Mode everywhere on the distribution's support. For example, the uniform distribution on a ball.
III. Modal-ridge along a circle. For example, the uniform distribution on a ring; see Example 1.

While the list is not exhaustive, the three types of modalities, referred to as Type I, II and III modalities, are most relevant to our discussion. In particular, the shape of the density function along any ray from the origin has only one peak, or one connected set of peaks, for these three types of modality. Let $f_{\mathbf{X}}$ be the density function of the rotationally symmetric distribution about the origin and suppose $\mathbf{X} = (X_1, \ldots, X_d)^\top \sim f_{\mathbf{X}}$. Then the section of $f_{\mathbf{X}}$ along any ray from the origin is proportional to the conditional density function of X_1 given $(X_2, \ldots, X_p) = \mathbf{0}$. That is, for any $\mathbf{v} \in \mathbb{R}^d$ with $\|\mathbf{v}\| = 1$,

$$f(x_1 \mid (X_2, \ldots, X_d) = \mathbf{0}) \;\propto\; f_{\mathbf{X}}(x_1\mathbf{v}) = f_{\mathbf{X}}((x_1, 0, \ldots, 0)). \tag{2}$$

Define $g(r) = f_{\mathbf{X}}((r, 0, \ldots, 0))$ for $r \geq 0$. It can be shown that $g(r)$ is proportional to $f_R(r)/r^{d-1}$, where f_R is the probability density function of $R = \|\mathbf{X}\|$. We have then

$$f_R(r) = \frac{1}{c_g(d-1)} r^{d-1} g(r),$$

where $c_g(p)$ is defined later in Lemma 2.

We write $f_{\mathbf{X}} = f_g$ as $f_{\mathbf{X}}$ is completely determined by g, given that $f_{\mathbf{X}}$ is rotationally symmetric. Any function $g : [0, \infty) \to [0, \infty)$ proportional to the radial section of $f_{\mathbf{X}}$ completely determines the type of modality. If $g(r)$ is strictly decreasing, then there is the unique mode at the origin. If $g(r)$ is constant on $[0, \theta)$, then $\mathbf{X}$ is uniform on the θ-ball centered at $\mathbf{0}$. More generally, if $g(r)$ is non-increasing, then the mode of f is either (uniquely) at $\mathbf{0}$ or at everywhere on the θ_g-ball

centered at $\mathbf{0}$. Here, $\theta_g = \sup_r\{r \geq 0 : g(r) = g(0)\}$. The condition of non-increasing g corresponds to Type I and II of the modalities, where the special case of constant g corresponds to Type II.

Alternatively, if, for a constant $c > 0$, g is non-decreasing on $[0, c)$ and non-increasing on $[c, \infty)$, then f has modes at a ring, $\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \in [\theta_1, \theta_2]\}$ where $\theta_1 \leq c \leq \theta_2$. This condition corresponds to Type III modality.

All other cases of modalities can be obtained by altering the shape of g. For example, if the curve has a number of risings and fallings, then $f_{\mathbf{X}}$ may have many rings of modes. We shall not consider these atypical cases.

A referee pointed out that in the literature of elliptical distributions the density generator g_U has been routinely used in analysis [19]. The density generator for the case of rotationally symmetric distribution is the density function of $U = \|\mathbf{X}\|^2$, and it does not render a simple connection to the types of modalities as the sectional density g does.

The multivariate kurtosis is easily computed from g:

Lemma 2. *For any rotationally symmetric density $f_{\mathbf{X}}$ on $\mathbb{R}^d$, let $g(r) = f_{\mathbf{X}}((r, 0, \ldots, 0))$ and $c_g(p) = \int_0^\infty r^p g(r)dr$ for $p \geq 0$. Then,*

$$\kappa(f_{\mathbf{X}}) = d^2 \frac{c_g(d+3)c_g(d-1)}{[c_g(d+1)]^2}. \tag{3}$$

Proof of Lemma 2. Suppose $\mathbf{X} \sim f_{\mathbf{X}}$. By Lemma 1(i), $\mathrm{Cov}(\mathbf{X}) = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_d$ for $\sigma^2 = d^{-1}\mathrm{trace}(\mathrm{Cov}(\mathbf{X})) = d^{-1}\mathrm{E}\|X\|^2$. Then $\kappa(f_{\mathbf{X}}) = \mathrm{E}(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1}\mathbf{X})^2 = d^2\mathrm{E}\|\mathbf{X}\|^4/(\mathrm{E}\|\mathbf{X}\|^2)^2$. Note that for any $m \geq 0$, $\mathrm{E}\|\mathbf{X}\|^m = \mathrm{E}R^m$, for $R \sim f_R$. Plugging in

$$\mathrm{E}R^m = \int_0^\infty r^m f_R(r)dr = \int_0^\infty \frac{1}{c_g(d-1)} r^{m+d-1} g(r)dr = \frac{c_g(d-1+m)}{c_g(d-1)}$$

and rearranging give the result. □

Example 1. For reference, we compute $\kappa(f_{\mathbf{X}})$ for a few examples. The multivariate kurtosis of the d-variate standard normal distribution is $d(d+2)$. There is only one family of distributions exhibiting Type II modality, which is the ball uniform distribution on $\mathbb{R}^d$. The kurtosis of the ball uniform distribution is

$$\kappa(\mathcal{U}_d) := \frac{d(d+2)^2}{d+4}. \tag{4}$$

For Type III modality, we consider a simple family of distributions. For $0 \leq \theta_1 < \theta_2 < \infty$, let f_{θ_1,θ_2} be the density function of the uniform distribution on the ring $\{\mathbf{x} \in \mathbb{R}^d : \theta_1 \leq \|\mathbf{x}\| \leq \theta_2\}$. Then the sectional density of f_{θ_1,θ_2} is proportional to $g(r) = 1_{\theta_1 \leq r \leq \theta_2}$. The multivariate kurtosis of f_{θ_1,θ_2} depends only on the parameter $\eta = \theta_1/\theta_2$, and is

$$\kappa(f_{\theta_1,\theta_2}) = \frac{d(d+2)^2}{d+4} \frac{(1-\eta^{d+4})(1-\eta^d)}{(1-\eta^{d+2})^2}.$$

Since $\partial\kappa(f_{\theta_1,\theta_2})/\partial\eta < 0$ for $\eta \in (0, 1)$, the multivariate kurtosis of the ring uniform distribution decreases as it moves away from the ball uniform. In particular, for $\eta = 0$, the corresponding density f_{0,θ_2} is that of the θ_2-ball uniform and $\kappa(f_{0,\theta_2}) = \kappa(\mathcal{U}_d)$. As $\eta \to 1$, $\kappa(f_{\eta\theta_2,\theta_2}) \to d^2$.

We now formally connect the multivariate kurtosis with the three types of modalities. Theorem 1 provides a sharp distinction of $\kappa(f_g)$ based on the shape of the sectional density function g.

Theorem 1. *Let $\mathbf{X} \sim f_g$, where the rotationally symmetric f_g is defined through a sectional density g. Assume that $\sup_r g(r) < \infty$.*

(i) *If g is non-increasing, then*

$$\kappa(f_g) \geq \frac{d(d+2)^2}{d+4}. \tag{5}$$

(ii) *If for a fixed $c > 0$, g is non-decreasing on $[0, c]$, and $g(r) = 0$ for $r > c$, then*

$$\kappa(f_g) \leq \frac{d(d+2)^2}{d+4}. \tag{6}$$

In both cases, the equality holds if and only if f_g is the density of a ball uniform distribution.

In our proof of Theorem 1(i), we use the fact that the sequence of Stieltjes moments is logarithmically convex. Here, Stieltjes moments are of the form $b_k = \int_0^\infty x^k d\mu(x)$, for any nonnegative measure μ. For the condition (ii), we use the well-known result that the sequence of Hausdorff moments, a special case of Stieltjes moments, is not only logarithmically convex, but also completely monotonic. Hausdorff moments are of the form $c_k = \int_0^1 x^k dF(x)$ for any probability measure F on the unit interval [36]. The condition of monotone g is essential for creating such measures. Decreasing densities on the positive real line have been studied for some time; see [2,33].

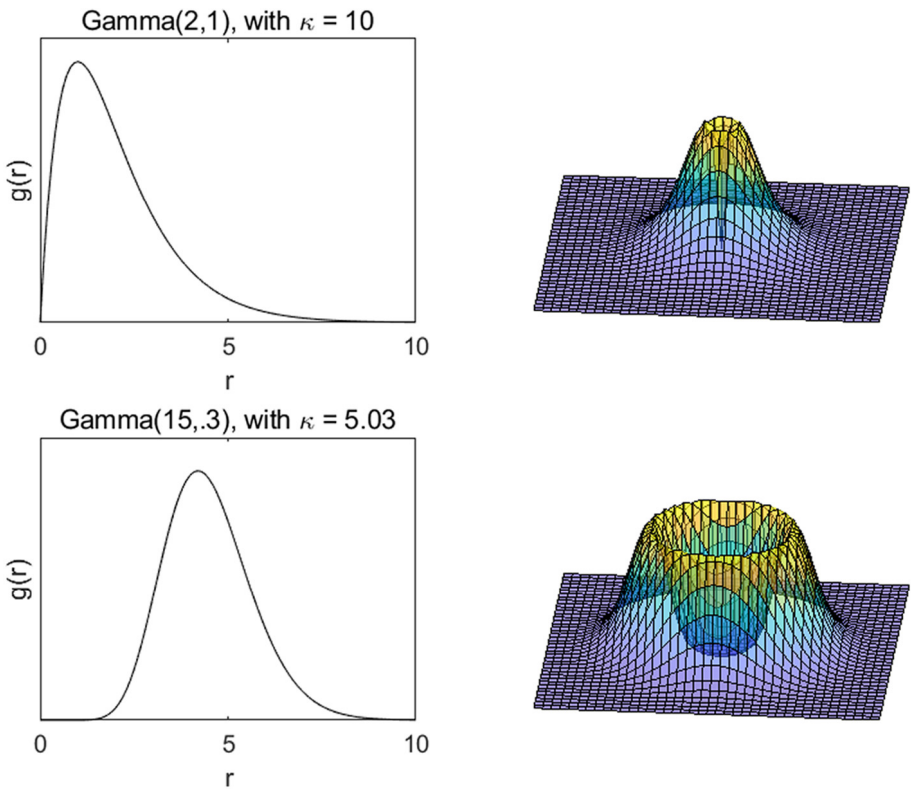

Fig. 2. Gamma densities as the sectional density g. Left panels show the shape of g. Right panels show the surface of f_g. While Gamma(2, 1) has a modal ridge, the overall shape of the distribution looks close to a rotationally symmetric unimodal distribution. The multivariate kurtosis $\kappa(\text{Gamma}(a, b))$ is compared with $\kappa(\mathcal{U}_2) = 5.3$.

Remark 1. The condition $\sup_r g(r) < \infty$ in Theorem 1 may be relaxed to allow $\lim_{r \searrow 0} g(r) = \infty$. A proof for such a case involves an approximation argument as used in [2]. However, in the interest of space and clarity, we prove our result only under the more restrictive assumption that we have stated.

Theorem 1 characterizes the three types of modalities via the multivariate kurtosis. In particular, the distribution with Type I modality, characterized by decreasing g, has $\kappa(f_g)$ greater than the kurtosis of the ball uniform. On the other hand, the distribution with increasing g (with a finite support), exhibiting Type III modality, has $\kappa(f_g)$ smaller than the kurtosis of the ball uniform. On the boundary is the ball uniform. This observation motivates our test procedure in Section 2.3.

We note that the shapes of g covered in the theorem are not exhaustive enough to cover all cases of Type III modality. A potentially more general family for Type III-exhibiting g includes the increasing-then-decreasing functions, as discussed earlier. One may wish that (6) holds under such an assumption. Unfortunately, it is not always true. As observed by [12] for the univariate case, distributions with a modal-ridge can have large kurtosis, which occurs when the modes of g are close to $\mathrm{E}\|\mathbf{X}\|^2$. An example of such a case is given by the g proportional to the gamma density with the shape parameter $\alpha > 1$. With the scale parameter $\beta = 1$, the resulting f_g has a modal-ridge with radius $\alpha - 1$, but $\kappa(f_g) = d^2(d + 2 + \alpha)(d + 1 + \alpha)/\{(d + \alpha)(d - 1 + \alpha)\} > \kappa(\mathcal{U}_d)$ for small α. (Note that $\kappa(f_g) < \kappa(\mathcal{U}_d)$ for large enough α.) This example tells us that even when the modal-ridge is present, if the modes are close to the center (or equivalently there exist excessively large masses at the tail), then the multivariate kurtosis is larger than that of the ball uniform. In Fig. 2, we show such an example; see the top panels. One may view such a modal-ridge distribution "close" to have a unimodality. On the other hand, when the basin at the center is wide enough (see, e.g., the bottom panels of Fig. 2), the corresponding kurtosis is smaller than $\kappa(\mathcal{U}_d)$.

To formalize the discussion above, we characterize the types of modality, precisely in this case, in terms of the coefficient of variation of the squared distance $\|\mathbf{X}\|^2$.

Theorem 2. *Suppose* $\mathbf{X} \sim f_{\mathbf{X}}$, *and* $f_{\mathbf{X}}$ *is rotationally symmetric about* $\mathbf{0}$. *Then* $\kappa(f_{\mathbf{X}}) < \kappa(\mathcal{U}_d)$ *if and only if*

$$\mathrm{c_v}(\|\mathbf{X}\|^2) := \frac{\mathrm{sd}(\|\mathbf{X}\|^2)}{\mathrm{E}\|\mathbf{X}\|^2} < \frac{2}{\sqrt{d(d+4)}}.$$

The statement is true when both inequalities are flipped, or converted to the equality.

Proof of Theorem 2. The result is easily derived from observing $d^{-2}\kappa(f_{\mathbf{X}}) = \mathrm{E}\|\mathbf{X}\|^4/(\mathrm{E}\|\mathbf{X}\|^2)^2 = \mathrm{Var}(\|\mathbf{X}\|^2)/(\mathrm{E}\|\mathbf{X}\|^2)^2 + 1$. □

The coefficient of variation is small when the mass of $\|\mathbf{X}\|^2$ is concentrated around its (nonzero) mean, which corresponds to a clear Type III modality of $f_{\mathbf{X}}$, which in turn, characterized by the smaller-than-uniform multivariate kurtosis.

There are mainly two causes for large values of $\mathrm{c_V}(\|\mathbf{X}\|^2)$. First and foremost, Theorem 1(i) shows that a monotone decreasing g leads to $\mathrm{c_V}(\|\mathbf{X}\|^2) > 2\{d(d+4)\}^{-1/2}$, as desired. Second, heavy tails of $\|\mathbf{X}\|$ (thus even heavier tails for $\|\mathbf{X}\|^2$) often lead to large $\mathrm{c_V}(\|\mathbf{X}\|^2)$. In view of this observation, the finite support condition of Theorem 1(ii) can be understood as one way of preventing heavy tails.

2.2. Modified sample kurtosis

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ be a random sample from the density $f_{\mathbf{X}}$. The sample multivariate kurtosis is defined as $\hat{k}(\mathbf{X}_1, \ldots, \mathbf{X}_n) = n^{-1}\sum_{i=1}^n\{(\mathbf{X}_i - \bar{\mathbf{X}})^\top \mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})\}^2$, where $\bar{\mathbf{X}}$ and $\mathbf{S}$ are the sample mean vector and covariance matrix computed from the sample [25]. When rotational symmetry is assumed, the covariance matrix is a scaled identity, and we replace $\mathbf{S}$ with the pooled sample variance $s^2 = \sum_{i=1}^n(\mathbf{X}_i - \bar{\mathbf{X}})^\top(\mathbf{X}_i - \bar{\mathbf{X}})/d(n-1)$ and define a modified sample kurtosis

$$\tilde{k}_n = \tilde{k}(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \frac{\sum_{i=1}^n \|\mathbf{X}_i - \bar{\mathbf{X}}\|^4/n}{\left\{\sum_{i=1}^n \|\mathbf{X}_i - \bar{\mathbf{X}}\|^2/d(n-1)\right\}^2}. \tag{7}$$

Under the rotational symmetry assumption, the modified sample kurtosis is consistent with the population kurtosis. In the following, we write $\mu_k = \mathrm{E}\|\mathbf{X}\|^k$, $\sigma_{ks} = \mathrm{Cov}(\|\mathbf{X}\|^k, \|\mathbf{X}\|^s)$, $k, s = 1, \ldots, 4$, for $\mathbf{X} \sim f_{\mathbf{X}}$.

Theorem 3. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$ be a random sample from the rotationally symmetric density $f_{\mathbf{X}}$. Suppose that the eighth moments of $\mathbf{X}_i$ are all finite. If $\mathrm{Var}(\mathbf{X}_1) = \sigma^2\mathbf{I}_d$ for some $\sigma^2 > 0$, then $\tilde{k}_n \to \kappa(f_{\mathbf{X}})$ in probability as $n \to \infty$, and*

$$\sqrt{n}(\tilde{k}_n - \kappa(f_{\mathbf{X}})) \to \mathcal{N}(0, \tau^2_{f_{\mathbf{X}}})$$

in distribution as $n \to \infty$, where

$$\tau^2_{f_{\mathbf{X}}} = d^4\left(\frac{\sigma_{44}}{\mu_2^4} - \frac{4\mu_4\sigma_{24}}{\mu_2^5} + \frac{4\mu_4^2\sigma_{22}}{\mu_2^6}\right). \tag{8}$$

Proof of Theorem 3. Without loss of generality, assume $\mathrm{E}\mathbf{X}_1 = \mathbf{0}$. The generous moment condition leads that $\|\bar{\mathbf{X}}\|^k = O_p(n^{-k/2})$ and $n^{-1}\|\mathbf{X}_i\|^k = O_p(1)$, for $k = 1, \ldots, 4$. Thus, by expanding the numerator and denominator of $\tilde{k}$ separately, it can be shown that

$$\tilde{k} = \frac{n^{-1}\sum_{i=1}^n \|\mathbf{X}_i\|^4 + O_p(n^{-1})}{d^{-2}\left(n^{-1}\sum_{i=1}^n \|\mathbf{X}_i\|^2\right)^2 + O_p(n^{-1})} = \frac{d^2\overline{R^4}}{(\overline{R^2})^2} + O_p(n^{-1}),$$

where $\overline{R^k} = n^{-1}\sum_{i=1}^n \|\mathbf{X}_i\|^k$. Again by the moment condition, $\tilde{k}$ converges in probability to $\kappa(f_{\mathbf{X}}) = d^2\mathrm{E}\|\mathbf{X}\|^4/(\mathrm{E}\|\mathbf{X}\|^2)^2$ as $n \to \infty$. The asymptotic normality is obtained by the central limit theorem and multivariate delta method. □

Note that the covariance condition of Theorem 3 is satisfied for any rotationally symmetric distribution. The moment condition might be more generous than needed. However, in our applications to directional data, all moments exist as the support of the distribution is bounded. See Section 3.

2.3. Test of modality

Our aim is to create a test of modality for rotationally symmetric multivariate distributions. In Section 2.1 we have observed that when the sectional density g is decreasing, Type I modality occurs, in which case $\kappa(f_{\mathbf{X}}) > \kappa(\mathcal{U}_d)$. Thus our hypotheses H_0 : "$f_{\mathbf{X}}$ exhibits Type I modality" versus H_1 : "$f_{\mathbf{X}}$ exhibits Type III modality" transform into

$$H_0 : \kappa(f_{\mathbf{X}}) \geq \kappa(\mathcal{U}_d) \text{ vs } H_1 : \kappa(f_{\mathbf{X}}) < \kappa(\mathcal{U}_d). \tag{9}$$

Note that the infimum of the kurtosis over non-increasing g is $\kappa(\mathcal{U}_d)$. So the kurtosis of a ball uniform is included in the null hypothesis.

For a significance level $\alpha > 0$, Theorem 3 provides an asymptotic level-α test procedure to test (9). In particular, to test a slightly general hypothesis $H_0' : \kappa(f_{\mathbf{X}}) \geq \kappa(f_0)$, for a prescribed f_0, using a random sample of size n, we reject H_0' when $\tilde{k}_n \leq \kappa(f_0) + z_\alpha\tau_{f_0}/\sqrt{n}$, where $z_\alpha = \Phi^{-1}(\alpha)$, and $\Phi(\cdot)$ is the standard normal distribution function.

For the test of modality, we take f_0 as a ball uniform. Then $\kappa(f_0) = \kappa(\mathcal{U}_d) = d(d+2)^2/(d+4)$ and $\tau^2_{\mathcal{U}_d} = 128d(d+2)^4/(d+4)^3(d+6)(d+8)$. The critical region of the modality test for hypotheses (9) is then

$$R_{\mathcal{U}} := \{k : k \leq \kappa(\mathcal{U}_d) + z_\alpha\tau_{\mathcal{U}_d}/\sqrt{n}\}. \tag{10}$$

The power function $\beta_n(f) = P_f(\tilde{k} \in R_{\mathcal{U}})$ is approximately

$$\beta_n(f_{\mathbf{X}}) \approx \Phi\left(\frac{\kappa(\mathcal{U}_d) - \kappa(f_{\mathbf{X}}) + z_\alpha \tau_{\mathcal{U}_d}/\sqrt{n}}{\tau_{f_{\mathbf{X}}}/\sqrt{n}}\right).$$

Thus, when $f_{\mathbf{X}}$ exhibits Type I modality, $\kappa(f_{\mathbf{X}}) > \kappa(\mathcal{U}_d)$ and the size of the test is approximately less than α. When $f_{\mathbf{X}}$ exhibits Type II modality (i.e. a ball uniform), then $\beta_n(f_{\mathbf{X}}) \approx \alpha$. It can be shown that the test is consistent as $n \to \infty$, i.e., $\lim_{n\to\infty} \beta_n(f_{\mathbf{X}}) = 0, \alpha$, or 1 when $\kappa(f_{\mathbf{X}})$ is greater than, equal to, or less than $\kappa(\mathcal{U}_d)$, respectively.

3. Method for directional data

3.1. Modalities of directional data

For $d \geq 2$, let $\mathbf{X} \in \mathbb{S}^d$ be a random direction whose distribution is given by the probability density function $f : \mathbb{S}^d \to [0, \infty)$ with respect to the Lebesgue measure on $\mathbb{S}^d$. Many parametric distributions on $\mathbb{S}^d$ are rotationally symmetric with respect to a location $\boldsymbol{\mu} \in \mathbb{S}^d$. As long as $f_{\mathbf{X}}$ depends only on $\boldsymbol{\mu}^\top \mathbf{x}$, it is rotationally symmetric. Examples include the von Mises–Fisher, Watson, and Bingham–Mardia distributions; see [26] and references therein. The von Mises–Fisher distribution, whose density is proportional to $\exp(c\boldsymbol{\mu}^\top \mathbf{x})$, is a normal-like distribution on $\mathbb{S}^d$ with the unique mode at $\boldsymbol{\mu}$, corresponding to Type I modality; see Fig. 1(a). On the other hand, the Bingham–Mardia distribution has a density proportional to $\exp\{c(\boldsymbol{\mu}^\top \mathbf{x} - \nu)^2\}$, and exhibits a modal ridge along a circle, that is, Type III modality; see Fig. 1(c) and (e).

For a distribution exhibiting Type II modality, we define a ball uniform distribution on $\mathbb{S}^d$. For $\tau \in (0, \pi]$, let $\mathcal{B}_\tau(\boldsymbol{\mu}) = \{\mathbf{x} \in \mathbb{S}^d : \mathbf{x}^\top \boldsymbol{\mu} \geq \cos(\tau)\}$ be the geodesic ball of angular radius τ, centered at $\boldsymbol{\mu} \in \mathbb{S}^d$. We will say that $\mathbf{X}$ has the spherical τ-ball uniform distribution if its density function $f_{\mathbf{X}}$ is proportional to $1_{\mathbf{x} \in \mathcal{B}_\tau(\boldsymbol{\mu})}$. To distinguish, we will refer to the usual ball uniform distribution on $\mathbb{R}^d$ as Euclidean ball uniform.

While it is tempting to consider the multivariate kurtosis $\kappa(f_{\mathbf{X}})$ as an indicator of the modality types, it is only defined for distributions on $\mathbb{R}^d$. Moreover, it is not clear what would be a generalization of the quadratic form $(\mathbf{x} - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ for directions. One may argue that $\mathbb{S}^d$ is indeed embedded in $\mathbb{R}^{d+1}$, but the random direction is no longer rotationally symmetric in $\mathbb{R}^{d+1}$ when it is so on $\mathbb{S}^d$. Furthermore, the density function is not even defined with respect to the Lebesgue measure on $\mathbb{R}^{d+1}$. To circumvent these issues, we use the tangent space approximation.

3.2. Spherical kurtosis for directional distributions

We first provide necessary notions. The tangent space $T_{\boldsymbol{\mu}}(\mathbb{S}^d) = \{\mathbf{z} \in \mathbb{R}^{d+1} : \mathbf{z}^\top \boldsymbol{\mu} = 0\}$ at $\boldsymbol{\mu} \in \mathbb{S}^d$ is the d-dimensional subspace of $\mathbb{R}^{d+1}$ whose elements are orthogonal to $\boldsymbol{\mu}$. We represent $T_{\boldsymbol{\mu}}(\mathbb{S}^d)$ by $T_{\boldsymbol{\mu}} \cong \mathbb{R}^d$, where $\mathbf{x} \in T_{\boldsymbol{\mu}}$ corresponds to $E_{\boldsymbol{\mu}}\mathbf{x} \in T_{\boldsymbol{\mu}}(\mathbb{S}^d)$. Here, the columns of the $(d+1) \times d$ matrix $E_{\boldsymbol{\mu}}$ are the basis of the null space of $\boldsymbol{\mu} \in \mathbb{R}^{d+1}$. The tangent space at $\boldsymbol{\mu}$ is used as a linear approximation of small neighbors of $\boldsymbol{\mu}$ on $\mathbb{S}^d$. Without loss of generality, assume that $\boldsymbol{\mu} = \mathbf{e}_1 := (1, 0, \ldots, 0)^\top \in \mathbb{S}^d$. The Exponential map at $\boldsymbol{\mu}$, $\mathrm{Exp}_{\boldsymbol{\mu}} : T_{\boldsymbol{\mu}} \to \mathbb{S}^d$, is defined as $\mathrm{Exp}_{\boldsymbol{\mu}}(\mathbf{y}) = (\cos\|\mathbf{y}\|, (\sin\|\mathbf{y}\|/\|\mathbf{y}\|)\mathbf{y}^\top)^\top$ for any $\mathbf{y} \in T_{\boldsymbol{\mu}}$. The inverse of Exponential map, called Log map for short, is defined on $\mathbb{S}^d$, minus the antipodal point of $\boldsymbol{\mu}$, and is $\mathrm{Log}_{\boldsymbol{\mu}} : \mathbb{S}^d \setminus \{-\boldsymbol{\mu}\} \to T_{\boldsymbol{\mu}}$,

$$\mathrm{Log}_{\boldsymbol{\mu}}((x_1, \ldots, x_{d+1})^\top) = \frac{\|\mathbf{x}\|}{\sin\|\mathbf{x}\|}(x_2, \ldots, x_{d+1})^\top.$$

The Exponential and Log maps preserve the distances to and angles at the point of tangency, and has been used to approximate the (empirical) distributions on $\mathbb{S}^d$ and on general manifolds; see, e.g., Ch. 13 of [26]. Finally, for $\mathbf{X} \sim f_{\mathbf{X}}$ on $\mathbb{S}^d$, the Fréchet mean of $\mathbf{X}$ is defined as the minimizer of the Fréchet function: $\boldsymbol{\mu}_{f_{\mathbf{X}}} = \mathrm{argmin}_{\boldsymbol{\mu} \in \mathbb{S}^d} \mathrm{E}\rho^2(\mathbf{X}, \boldsymbol{\mu})$, where $\rho(\mathbf{x}, \boldsymbol{\mu}) = \cos^{-1}(\mathbf{x}^\top \boldsymbol{\mu})$ is the geodesic (or intrinsic) distance function on $\mathbb{S}^d$. The Fréchet mean, sometimes called the geodesic mean, always exists on $\mathbb{S}^d$ but may not be unique.

We define the kurtosis of a directional distribution as the multivariate kurtosis of the Log-mapped distribution.

Definition 1 (*Spherical Kurtosis of Directional Distributions*)**.**

(i) Suppose that $\mathbf{X} \sim f$ on $\mathbb{S}^d$ has a unique Fréchet mean at $\boldsymbol{\mu} \in \mathbb{S}^d$. The kurtosis of $f_{\mathbf{X}}$ is defined as $\kappa(f_{\mathbf{X}}) := \mathrm{E}(\mathbf{Y}^\top \boldsymbol{\Sigma}^{-1}\mathbf{Y})^2$, where $\mathbf{Y} = \mathrm{Log}_{\boldsymbol{\mu}}(\mathbf{X})$ and $\boldsymbol{\Sigma} = \mathrm{E}(\mathbf{Y}\mathbf{Y}^\top)$.

(ii) Suppose that a sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathbb{S}^d$ has a unique solution $\hat{\boldsymbol{\mu}}$ of the empirical Fréchet problem $\min_{\boldsymbol{\mu} \in \mathbb{S}^d} \sum_{i=1}^n \rho^2(\mathbf{x}_i, \boldsymbol{\mu})$. The sample kurtosis of the sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is defined as $\hat{k} := \sum_{i=1}^n (\mathbf{y}_i^\top \mathbf{S}^{-1}\mathbf{y}_i)^2/n$, where $\mathbf{y}_i = \mathrm{Log}_{\hat{\boldsymbol{\mu}}}(\mathbf{x}_i)$ and $\mathbf{S}$ is the sample covariance matrix computed from $\mathbf{y}_i$'s.

In the above definition, the mean of $\mathbf{Y}$ is $\mathbf{0}$, due to the fact that the Fréchet mean is the point of tangency.

The necessary and sufficient condition for the density $f_{\mathbf{X}}$ being rotationally symmetric about $\boldsymbol{\mu}$ is that there is a function g_S such that $f_{\mathbf{X}}(\mathbf{x}) = g_S\{\cos^{-1}(\mathbf{x}^\top \boldsymbol{\mu})\}$. We write $f_{\mathbf{X}} = f_{g_S}$ for such a case. Moreover, the Fréchet mean of $f_{\mathbf{X}}$, if it is unique, is always $\boldsymbol{\mu}$.

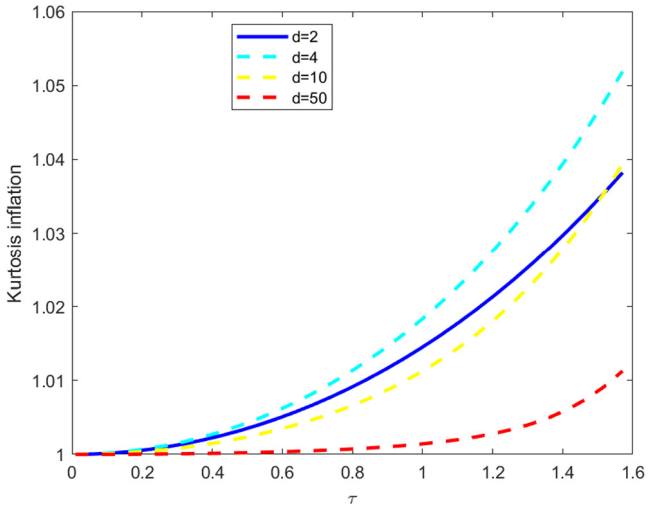

Fig. 3. The inflation factor $\kappa(\mathcal{S}_\tau)/\kappa(\mathcal{U}_d)$ measures the degree of inflation of the spherical kurtosis of spherical ball uniform in $\mathbb{S}^d$, compared with the multivariate kurtosis of Euclidean ball uniform.

The function g_S on $[0,\pi]$ has the same role as the sectional density g in the real vector space. The three types of modality described for the Euclidean case apply to directional distributions as well. That is, a non-increasing g_S leads to Type I or II modality. For instances, the unimodal von Mises distribution has a decreasing sectional density $\log g_S(r) \propto \cos(r)$, while for the spherical ball uniform, $g_S(r) \propto 1_{r\le\tau}$ is non-increasing. An increasing-then-decreasing shape of g_S corresponds to Type III. The Bingham–Mardia distribution has a circular mode and $\log g_S(r) \propto (\cos(r)-\nu)^2$.

Lemma 3. *Suppose that* $\mathbf{X} \sim f_\mathbf{X}$ *on* $\mathbb{S}^d$ *is rotationally symmetric about* $\boldsymbol{\mu}$ *and* $g_S : [0,\pi] \to [0,\infty)$ *satisfies* $f_\mathbf{X}(\mathbf{x}) = g_S(\cos^{-1}(\mathbf{x}^\top\boldsymbol{\mu}))$ *for all* $\mathbf{x} \in \mathbb{S}^d$. *Let* $C(m,d;g_S) = \int_0^\pi g_S(\phi)\phi^m(\sin\phi)^{d-1}d\phi$ *for* $m, d \in \{0,1,\ldots\}$. *Then*

(i) *the density function of* $\mathbf{Y} = \mathrm{Log}_{\boldsymbol{\mu}}(\mathbf{X})$ *is*

$$f_\mathbf{Y}(\mathbf{y}) = g_S(\|\mathbf{y}\|)\left(\frac{\sin\|\mathbf{y}\|}{\|\mathbf{y}\|}\right)^{d-1}, \text{ for } \mathbf{y} : \|\mathbf{y}\| \le \pi;$$

(ii) *The density function of* $\|\mathbf{Y}\|$ *is*

$$f_{\|\mathbf{Y}\|}(r) = \frac{g_S(r)\sin^{d-1}(r)}{C(0,d;g_S)}, \ r \in (0,\pi];$$

(iii) *The spherical kurtosis of* $f_\mathbf{X}$ *is*

$$\kappa(f_\mathbf{X}) = d^2\frac{C(4,d;g_S)C(0,d;d_S)}{[C(2,d;g_S)]^2}.$$

Lemma 3 paves a way to convert the problem of modality (through the sectional density g_S) into a problem of spherical kurtosis. However, unfortunately, a constant $g_S(r)$, corresponding to the spherical τ-ball uniform and Type II modality, does not lead to a constant $f_\mathbf{Y}(\mathbf{y})$. Moreover, the kurtosis of the spherical ball uniform depends on τ, as opposed to the transformation invariant $\kappa(\mathcal{U}_d)$ in the Euclidean case. This issue is demonstrated in the next example.

Example 2. Denoting the spherical τ-ball uniform distribution as $\mathcal{S}_\tau$, the kurtosis $\kappa(\mathcal{S}_\tau)$ is given by Lemma 3(iii) with $g_S(r) = 1\{r \le \tau\}$ and $C(m,d;g_S) = \int_0^\tau x^m(\sin x)^{d-1}dx$.

The precise closed form of $\kappa(\mathcal{S}_\tau)$ is not so informative, but an approximation of $\kappa(\mathcal{S}_\tau)$ sheds some light on the dependence of $\kappa(\mathcal{S}_\tau)$ on τ. In particular, the second-order polynomial approximation is

$$\check{\kappa}(\mathcal{S}_\tau) = \kappa(\mathcal{U}_d)\left(1 + c_d\tau^2\right) \approx \kappa(\mathcal{S}_\tau), \quad c_d = \frac{8(d-1)}{3(d+2)(d+4)(d+6)}.$$

This approximation is more accurate if τ is close to 0. It can be checked that $\check{\kappa}(\mathcal{S}_0) = \lim_{\tau\to 0}\kappa(\mathcal{S}_\tau) = \kappa(\mathcal{U}_d)$, and, as τ increases, $\check{\kappa}(\mathcal{S}_\tau)$ increases but at a slow rate c_d. As shown in Fig. 3, the inflation of kurtosis, compared to $\kappa(\mathcal{U}_d)$, is not substantial.

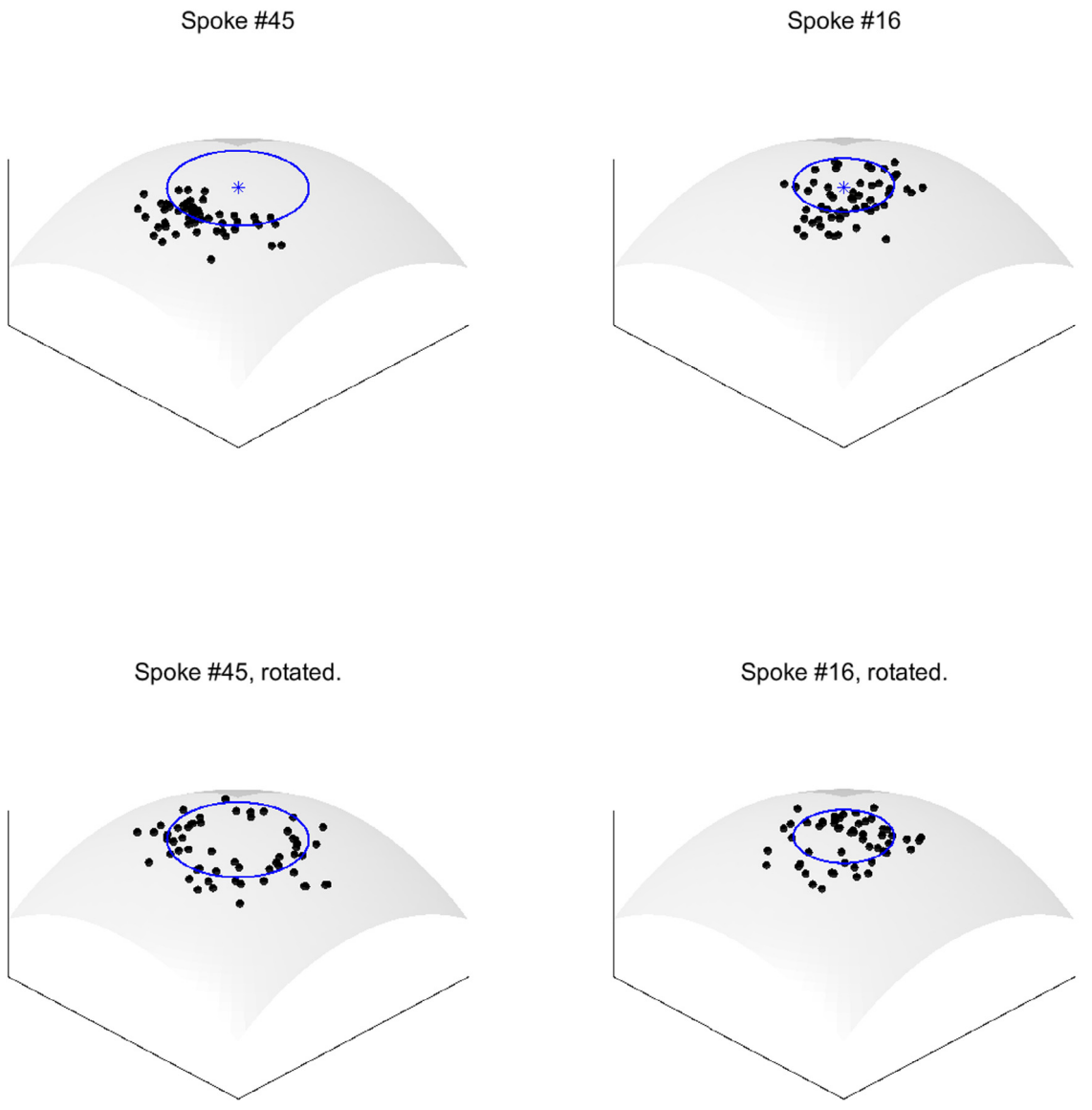

Fig. 4. Two sets of directions, called spoke directions, from the control group of skeletally represented hippocampus shape data. Top row shows the raw data (black dots) with the fitted small circle $C(\hat{\mathbf{c}}, \hat{r})$. The symbol '*' represents the center of the circle, $\hat{\mathbf{c}}$. Bottom rows display the randomly-rotated data $\mathbf{O}_i\mathbf{x}_i$, that are now rotationally symmetric about $\hat{\mathbf{c}}$.

3.3. Modified sample kurtosis

Suppose that for a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim f$ on $\mathbb{S}^d$, the sample Fréchet mean $\hat{\boldsymbol{\mu}} = \operatorname{argmin}_{\boldsymbol{\mu}\in\mathbb{S}^d} \sum_{i=1}^n \rho^2(\mathbf{X}_i, \boldsymbol{\mu})$ is unique. If, in addition, $f_{\mathbf{X}}$ is rotationally symmetric about $\boldsymbol{\mu}$, then a modified sample kurtosis is defined as

$$\tilde{k}_n = \tilde{k}(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \frac{\sum_{i=1}^n \|\mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}_i\|^4/n}{\left\{\sum_{i=1}^n \|\mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}_i\|^2/d(n-1)\right\}^2}. \tag{11}$$

We show in Theorem 4 that $\tilde{k}_n$ is a consistent estimator of $\kappa(f_{\mathbf{X}})$ and exhibits an asymptotic normality, under the following conditions:

Assumption 1. The sample Fréchet mean $\hat{\boldsymbol{\mu}}$ is unique for all $n > N$ for some $N < \infty$.

Assumption 2. $\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = O_p(n^{-1/2})$.

These assumptions are not restrictive. Assumption 1 is satisfied if the support of the distribution is small enough or with a high probability if the population Fréchet mean is unique, which is true for most rotationally symmetric distributions. Assumption 2 is satisfied when $\hat{\boldsymbol{\mu}}$ is $\sqrt{n}$-consistent. For example, Theorem 2.1 of [3] states that $\sqrt{n}\mathrm{Log}_{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}$ converges to a non-degenerate mean-zero normal distribution, under regularity conditions. Assumption 2 is then satisfied since $\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = \|\mathrm{Log}_{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}\|$.

Theorem 4. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{S}^d$ are i.i.d. with density $f_{\mathbf{X}}$. Suppose that $f_{\mathbf{X}}$ is rotationally symmetric about $\boldsymbol{\mu} \in \mathbb{S}^d$ and its Fréchet mean is $\boldsymbol{\mu}$, and Assumptions 1 and 2 are satisfied. Then, $\tilde{k}_n \to \kappa(f_{\mathbf{X}})$ in probability as $n \to \infty$, and $\sqrt{n}(\tilde{k}_n - \kappa(f_{\mathbf{X}})) \to N(0, \tau^2_{f_{\mathbf{X}}})$ in distribution as $n \to \infty$, where $\tau^2_{f_{\mathbf{X}}}$ is given in (8), with $\mu_k = \mathrm{E}\|\mathbf{Y}\|^k$, $\sigma_{ks} = \mathrm{Cov}(\|\mathbf{Y}\|^k, \|\mathbf{Y}\|^s)$, $k, s \in \{1, \ldots, 4\}$, for $\mathbf{Y} = \mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X}$, $\mathbf{X} \sim f_{\mathbf{X}}$.*

A key step in verifying Theorem 4 is to replace the sample Fréchet mean in $\tilde{k}_n$ with the population Fréchet mean, for which the following lemma is instrumental.

Lemma 4. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{S}^d$ are i.i.d. with $f_{\mathbf{X}}$. Suppose that Assumptions 1 and 2 are satisfied. Then $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}\| = O(n^{-1/2})$. Here, $\mathbf{X}$ may or may not be independent of $\hat{\boldsymbol{\mu}}$.*

While it is natural to expect the conclusion of Lemma 4, it is quite tricky to verify. For example, when $\mathbf{x}$ is close to the antipodal point of $\boldsymbol{\mu}$, then even if $\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) < \epsilon$, the tangent vectors $\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x}$ and $\mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{x}$ may point opposite directions. For example, let $\boldsymbol{\mu} = (1, 0, 0)^\top$, $\hat{\boldsymbol{\mu}} = (\cos(t), -\sin(t), 0)^\top$ and $\mathbf{x} = (-\cos(t/2), \sin(t/2), 0)^\top$ on $\mathbb{S}^2$. As $t \to 0$, $\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = t \to 0$, but $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{x}\| \to \pi$. In our proof of Lemma 4, the assumption of absolute continuity of the random variable $\rho(\boldsymbol{\mu}, \mathbf{X})$ is used to bound the probability of non-degenerating $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}\|$.

Proof of Theorem 4. The proof is similar to the proof of Theorem 3. By Lemma 4, we get $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X}_i - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}_i\|^k = O_p(n^{-k/2})$. By expanding the numerator and denominator of $\tilde{k}_n$ (11) separately, we get

$$\tilde{k}_n = d^2 \frac{n^{-1}\sum_{i=1}^n \|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X}_i\|^4 + O_p(n^{-1})}{\left(n^{-1}\sum_{i=1}^n \|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X}_i\|^2\right)^2 + O_p(n^{-1})} = \frac{d^2\overline{R^4}}{(\overline{R^2})^2} + O_p(n^{-1}),$$

where $\overline{R^k} = n^{-1}\sum_{i=1}^n \|\mathbf{Y}_i\|^k$. The rest of proof is identical to Theorem 3, with $\mathbf{X}_i$ replaced by $\mathbf{Y}_i$. □

3.4. Test of modality

As done for the Euclidean case in Section 2.3, Type I and II modalities are characterized by the non-increasing sectional density g_S for the directional distributions. Write f_{g_S} for the corresponding rotationally symmetric density function on $\mathbb{S}^d$. If g_S is non-increasing, then $g(r) = (\sin r/r)^{d-1} g_S(r)$ is strictly decreasing for $d \geq 2$, and by Theorem 1, $\kappa(f_{g_S}) > \kappa(\mathcal{U}_d)$. This strict inequality is true for f_{g_S} being the density of a spherical ball uniform, for $d \geq 2$. To represent the null hypothesis of unimodality as the form of $\kappa(f_{\mathbf{X}}) \geq \kappa^\circ$, we set κ° as the infimum kurtosis over the set of distributions with non-increasing g_S. In fact,

$$\kappa(\mathcal{U}_d) = \inf_{g_S \in \mathcal{G}^-} \kappa(f_{g_S}),$$

where $\mathcal{G}^-$ is the set of all non-increasing functions on $[0, \pi)$.

Thus, in the unimodality test for rotationally symmetric directional distributions against Type III modality, the hypotheses are exactly the same as in (9). The test procedure at level α is exactly the same as in Section 2.3, except that the test statistic $\tilde{k}_n$ is computed by (11). The theoretical properties of the modality test, shown in Section 2.3, apply here as well due to Theorem 4.

4. Application to dimension reduction of directional data

4.1. Overfit of principal nested spheres

The analysis of principal nested spheres (PNS) [15] is a principal component analysis analog for directional data on $\mathbb{S}^d$. It is an example of backward dimension reduction [5]: A $(p-1)$-dimensional submanifold, containing the largest total variance, is fitted in $\mathbb{S}^p$, successively for decreasing $p \in \{d, d-1, \ldots, 2\}$. We restrict our attention to $\mathbb{S}^2$ for clarity, but our discussion applies to higher dimensions.

For data on $\mathbb{S}^2$, Jung et al. [15] chose to use circles as candidates for the 1-dimensional submanifold. A circle $C(\mathbf{c}, r) = \{\mathbf{x} \in \mathbb{S}^2 : \rho(\mathbf{x}, \mathbf{c}) = r\}$ with spherical radius $r = \pi/2$ corresponds to a geodesic dimension reduction while a choice of $r < \pi/2$ leads to a non-geodesic dimension reduction, analogous to linear or non-linear dimension reduction in Euclidean

spaces. Here, the circle $C(\mathbf{c}, r)$ can be thought of as an analog to the dimension-1 subspace spanned by the first principal component direction in $\mathbb{R}^2$. Jung et al. proposed to solve the least-square problem of minimizing

$$\sum_{i=1}^{n}\{\rho(\mathbf{c}, \mathbf{x}_i) - r\}^2. \tag{12}$$

An important question is whether it is advantageous to allow the non-geodesic submanifold fitting, given by $r < \pi/2$. Sometimes, a non-geodesic dimension reduction results in a severe overfitting of the data. For example, fitting $C(\mathbf{c}, r)$ by (12) to the data examples in Fig. 1(a)–(a_3) results in a small circle with unnecessarily small r, an overfit of the data. The problem of overfitting resides also for the data examples in Fig. 1(b) and (c), where the fitted radius $\hat{r}$, solving (12), is less than $\pi/2$ with probability 1. A non-geodesic dimension reduction is appropriate for cases (d) and (e).

Jung et al. [15] proposed to use a two-step procedure to decide a geodesic or non-geodesic fit:

Step 1. A goodness-of-fit test based on signed residuals, $\rho(\mathbf{x}_i, \hat{\mathbf{c}}) - \hat{r}$, is applied. If the null hypothesis of $r = \pi/2$ is not rejected, then a geodesic ($\hat{r} = \pi/2$) fit is chosen (corresponding to the cases (b) and (c) in Fig. 1), and Step 2 is skipped. If the best fitting r is significantly smaller than $\pi/2$ (corresponding the cases (a), (d) and (e)), then Step 2 is applied.

Step 2. A parametric bootstrap test with the null hypothesis of von Mises–Fisher distribution is applied to further discern the isotropic case (a) from (d) and (e).

While this procedure has been deemed useful [6,13,32,38], it comes with drawbacks. For a large sample size and high dimensions, the computation cost is very high. Moreover, when the underlying distribution is a spherical ball uniform, the parametric bootstrap test mistakenly suggests a non-geodesic fit.

4.2. Proposed method

We propose to replace the parametric bootstrap test in Step 2 by the test of modality developed in Section 3. The goodness of fit test in Step 1 is still used to discern the cases where great circle fit is clearly advantageous from the others. To better motivate our application of the kurtosis test, we use two sets of directions on $\mathbb{S}^2$ taken from real data, further discussed in Section 5.1. The top row of Fig. 4 displays these two sets of data, with fitted small circles. The small circle fit is deemed correct for the left, while it is an overfit for the right. Notice that the fitted circle $C(\hat{\mathbf{c}}, \hat{r})$ is desirable when there is no mass near the center $\hat{\mathbf{c}}$. On the other hand, it is overfitted when there are excess mass at $\hat{\mathbf{c}}$.

Let $\mathbf{O}_i$ be independently sampled from the uniform distribution on the set of orthogonal matrices satisfying $\mathbf{O}_i\hat{\mathbf{c}} = \hat{\mathbf{c}}$. For the raw data $\mathbf{x}_i$, define $\mathbf{z}_i = \mathbf{O}_i\mathbf{x}_i$. It can be checked that the distribution of $\mathbf{z}_i$ is rotationally symmetric about $\hat{\mathbf{c}}$, and the distance to the center is preserved, i.e. $\rho(\hat{\mathbf{c}}, \mathbf{z}_i) = \rho(\hat{\mathbf{c}}, \mathbf{x}_i)$. The latter implies that the mass of $\mathbf{z}_i$ near $\hat{\mathbf{c}}$ is equal to that of $\mathbf{x}_i$. The bottom row of Fig. 4 displays an example of rotated data. Type III modality is observed for the left, while Type I modality is observed for the right.

The hypothesis that the given small circle is an overfit is thus equivalent to the hypothesis that $\mathbf{z}_i$ has Type I modality, which can be tested by the kurtosis test. In practice, since $\|\mathrm{Log}_{\hat{\mathbf{c}}}\mathbf{z}_i\| = \|\mathrm{Log}_{\hat{\mathbf{c}}}\mathbf{x}_i\|$, there is no need to sample $\mathbf{O}_i$. The test statistic is

$$\tilde{k}_n = \frac{\sum_{i=1}^{n}\|\mathrm{Log}_{\hat{\mathbf{c}}}\mathbf{z}_i\|^4/n}{\left\{\sum_{i=1}^{n}\|\mathrm{Log}_{\hat{\mathbf{c}}}\mathbf{z}_i\|^2/d(n-1)\right\}^2} = \frac{\sum_{i=1}^{n}\|\mathrm{Log}_{\hat{\mathbf{c}}}\mathbf{x}_i\|^4/n}{\left\{\sum_{i=1}^{n}\|\mathrm{Log}_{\hat{\mathbf{c}}}\mathbf{x}_i\|^2/d(n-1)\right\}^2},$$

and if $\tilde{k}_n \in R_{\mathcal{U}}$ (10), the null hypothesis of unimodality is rejected, and the small circle fit is used. Otherwise, the small circle fit is discarded, and $C(\mathbf{c}, r)$ is fitted with the restriction $r = \pi/2$.

The proposed method works for $\mathbb{S}^d$, $d \geq 2$, as well.

4.3. Numerical comparison

We numerically compare the performance of the proposed procedure with the original procedure of [15] in Section 4.1. In addition, the Bayesian information criterion (BIC)-based decision rule, also discussed in [15], and the test proposed by [8] are compared. The likelihood ratio-based test of [8] assumes the folded normal distribution for residuals. Furthermore, the folded normal test is also used to replace the bootstrap test in Step 2 of the original procedure.

We consider several data situations in $\mathbb{S}^2$. For the rotationally symmetric cases, we use (a) a von Mises–Fisher, (a_1) a tangent normal, $\mathcal{N}_d(\mathbf{0}, \sigma^2\mathbf{I}_d)$ on $T_{\boldsymbol{\mu}}\mathbb{S}^2$, mapped to $\mathbb{S}^2$ by Exponential map, (a_2) a Euclidean ball uniform on $T_{\boldsymbol{\mu}}\mathbb{S}^2$, mapped to $\mathbb{S}^2$, and (a_3) a spherical ball uniform. We prefer geodesic dimension reduction for these cases. The cases where there is a clear mode of variation along a geodesic are represented by (b) a small sphere distribution of [17] with $r = \pi/2$ and (c) a Bingham–Mardia distribution with $r = \pi/2$ [4]. For the case where non-geodesic dimension reduction is preferred, we use (d) a small sphere distribution with $r < \pi/2$ and (e) a Bingham–Mardia with $r < \pi/2$. Fig. 1 displays example data sampled from these distributions.

As a measure of performance of the competing methods, we compute the proportion of correct decisions among 1000 repetitions. The decision of geodesic fit is correct for cases (a) to (c), while it is not for (d) and (e). In Table 1 we collect

Table 1
The proportion of correct decisions among geodesic and non-geodesic dimension reduction. The proposed method shows an overall good performance across all cases. See text for description of the cases (a)–(e). Methods are (1) Jung et al. [15], (2) folded normal test in [8], (3) the goodness of fit (Step 1) and Folded normal test (Step 2), sequentially applied, and (4) our proposal.

n	Method	(a)	(a_1)	(a_2)	(a_3)	(b)	(c)	(d)	(e)
50	Original[1]	0.93	0.94	0.09	0.10	0.93	0.95	0.99	1.00
	BIC[1]	0.00	0.00	0.00	0.00	0.93	0.95	0.99	1.00
	FN[2]	1.00	1.00	0.79	0.82	0.00	0.00	1.00	1.00
	GF + FN[3]	1.00	1.00	0.79	0.82	0.93	0.95	0.99	1.00
	Kurtosis[4]	1.00	1.00	0.80	0.83	0.93	0.95	0.99	1.00
200	Original	0.94	0.89	0.00	0.00	0.94	0.95	1.00	1.00
	BIC	0.00	0.00	0.00	0.00	0.97	0.98	1.00	1.00
	FN	1.00	1.00	0.70	0.71	0.00	0.00	1.00	1.00
	GF + FN	1.00	1.00	0.70	0.71	0.94	0.95	1.00	1.00
	Kurtosis	1.00	1.00	0.89	0.92	0.94	0.95	1.00	1.00

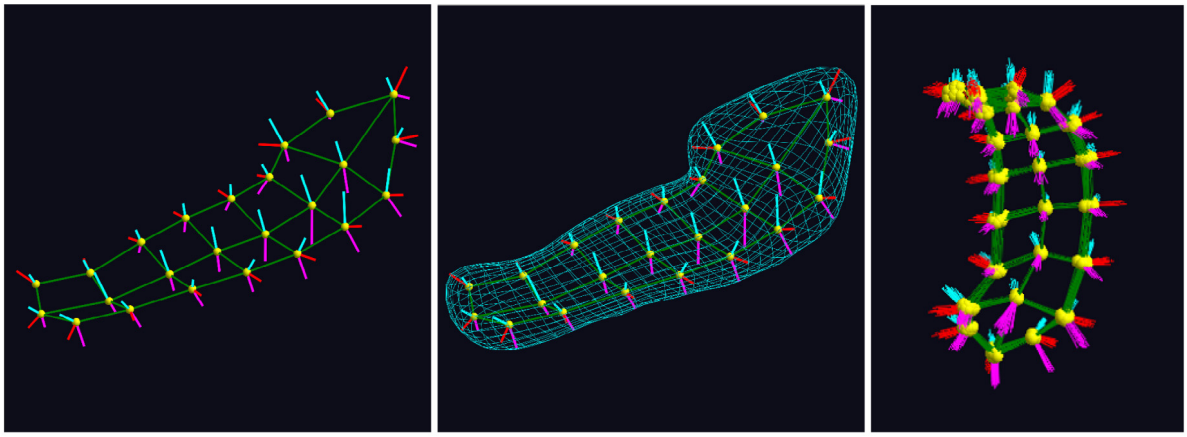

Fig. 5. (a) Skeletal representation (s-rep) model of a hippocampus. (b) S-rep with interpolated boundary. (c) Aligned and overlaid s-reps of 56 hippocampi in the control group.

the proportions computed for the sample size $n = 50, 200$. Overall, the correct decision rates of the proposed method are always among the highest, for all cases. The original method of [15] fails for uniform null distributions, in cases (a_2)–(a_3). The BIC procedure always fits small circles for rotationally symmetric cases (a)–(a_3). The method of [8] alone was not successful at all for cases (b) and (c), but the modified procedure (GF+FN in Table 1), with the goodness-of-fit test, solves the problem. For the ball uniform distributions, in cases (a_2)–(a_3), the folded normal test makes correct decisions for about 70% of the time, while our method shows a significantly superior accuracy of about 90%, for $n = 200$.

The computation times of the proposed method are significantly shorter than the times needed for the methods of [15] and [8].

5. Real data examples

5.1. Dimension reduction for hippocampus shapes

In medical imaging, it is of interest to characterize the major variation in 3D models of human brain. The PNS analysis has been used for the purpose [32,34]. We compare the proposed procedure in Section 4.2 with the existing options (described in Section 4.3) in PNS analysis of 3D shape representations of hippocampus in human brain.

The data consist of hippocampi from 221 patients with schizophrenia and 56 healthy controls [27,35]. For 3D modeling, skeletal representations (s-reps) were fitted to the data as detailed in [34]. In particular, a discrete slabular s-rep, with the 3×8 grid of skeletal positions $\mathbf{p}_i \in \mathbb{R}^3$, $i \in \{1, \ldots, 24\}$, is fitted to each hippocampus. In Fig. 5, $\mathbf{p}_i$ are shown as yellow balls. At each skeletal position, spoke-like vectors from $\mathbf{p}_i$ to the boundary of the hippocampus are fitted as well. These spoke vectors are decomposed into the length $r > 0$ and the direction $\mathbf{x} \in \mathbb{S}^2$. As a result, each s-rep is described by a tuple $\mathbf{s} = (\mathbf{p}_1, \ldots, \mathbf{p}_{24}, r_1, \ldots, r_{66}, \mathbf{x}_1, \ldots, \mathbf{x}_{66}) \in \mathbb{R}^{72} \times \mathbb{R}_+^{66} \times (\mathbb{S}^2)^{66}$. The shapes of skeletal positions $(\mathbf{p}_1, \ldots, \mathbf{p}_{24})$ are considered to lie in the preshape space, which is the hypersphere $\mathbb{S}^{72-4}$ [7]. The combined directional vector in $\mathbb{S}^{68}$ of

Table 2
Performances of testing procedures in the dimension reduction of hippocampi s-reps, for control ($n = 56$) and schizophrenic ($n = 221$) groups. The explained amount of variation and accuracy are shown for the number of components $m = 1, 5, 15$. The mean (standard deviation) of computation times from 25 repetitions are shown. Testing procedures are the same as in Table 1.

# components		% of variance			Approx. error			Comp. time
		1	5	15	1	5	15	(in s)
Controls	Original	32.71	75.57	93.60	30.81	18.28	8.93	185.72 (2.33)
	BIC	27.70	66.13	86.47	34.70	21.97	14.48	5.91 (0.16)
	FN	33.07	76.52	94.28	30.71	18.04	8.54	23.89 (0.23)
	GF + FN	33.16	76.55	94.40	30.70	18.03	8.45	13.32 (0.03)
	Kurtosis	33.16	76.55	94.40	30.70	18.03	8.45	5.40 (0.02)
Schizoph.	Original	40.07	76.95	92.89	64.19	39.39	20.88	496.84 (5.38)
	BIC	30.45	63.95	81.73	72.11	50.85	35.32	7.56 (0.08)
	FN	40.07	76.95	92.90	64.18	39.36	20.86	30.35 (0.05)
	GF + FN	40.07	76.95	92.89	64.19	39.39	20.88	20.56 (0.03)
	Kurtosis	40.07	76.95	92.89	64.19	39.39	20.88	7.66 (0.09)

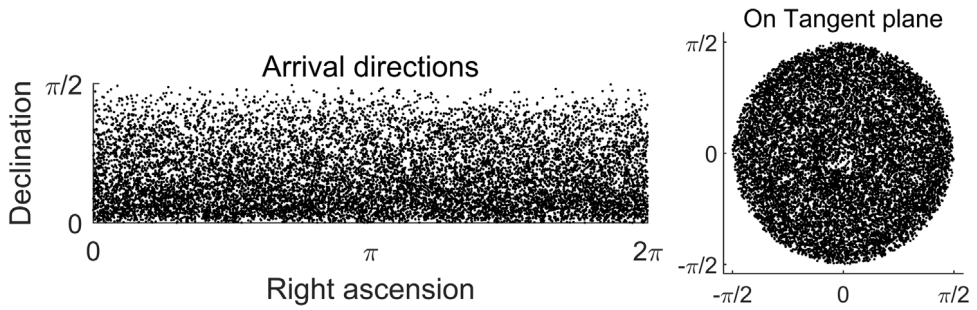

Fig. 6. (left) The scatter of arrival directions of $n = 12{,}877$ cosmic events, shown in terms of right ascension (longitude) and declination (latitude). (right) The scatter of $\mathbf{y}_i = \mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x}_i$, where $\boldsymbol{\mu}$ is set to the north pole.

skeletal positions and the spoke directions in $\mathbb{S}^2$ are each analyzed by PNS, the results of which are combined as proposed in [32]. As a result, we obtain a PCA-like decomposition of s-reps.

To compare the performance of the proposed method of preventing the overfit for this s-rep data, we computed the cumulative sum of variances contained in the first m principal components, and the accuracy of reconstructed s-reps from the first m components. The accuracy is measured by an approximation error, $\sum_{i=1}^{n} \|w_i - \tilde{w}_i(m)\|$, where w_i is the boundary positions (end points of the spokes) of the ith original s-rep, and $\tilde{w}_i(m)$ is the boundary positions of the ith reconstructed s-rep using the first m components. In addition, the computation times are recorded.

Results are reported in Table 2. The quality of dimension reduction using the proposed procedure is nearly identical to other reasonable options, if not deemed better. To investigate the difference between the original method and our proposal in the control group, we take two spoke directions in spoke #45 and #16; see Fig. 4. All testing procedures result in the small circle fit for spoke #45, as desired. The original procedure recommends a small circle fit for spoke #16, which is an overfit. The kurtosis test of modality does not reject the hypothesis of unimodality, resulting in a great circle fit, and in turn, a better dimension reduction.

The proposed method clearly outperforms the other methods in terms of computation time. When comparing two or more populations of s-reps, it is common to repeat the PNS analysis multiple times in, e.g., resampling-based comparisons. The significant reduction of computation times is beneficial in such studies.

5.2. Arrival directions of cosmic events

In astrophysics, the locations of cosmic events are often recorded as arrival directions on the celestial sphere. We use a record of muon neutrino candidate events recorded by the IceCube Neutrino Observatory in the 40 string configuration, during April 2008 to May 2009 [1]. From the data set containing $n = 12{,}877$ cosmic events, available at https://icecube.wisc.edu/science/data/ic40, we only keep the arrival directions, and demonstrate an application of the proposed test of modality. Fig. 6 displays the raw data. Since the observations are scattered across the northern celestial hemisphere, we assume that the data are rotationally symmetric about the north celestial pole. The Log-mapped data, at the north pole, are also shown in Fig. 6.

Since the Fréchet mean is assumed to be known as the north pole, we compute the sample kurtosis by $\tilde{k} = d^2 n \sum_{i=1}^{n} \|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x}_i\|^4 / \{\sum_{i=1}^{n} \|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x}_i\|^2\}^2 = 5.2703$. This is quite close to $\kappa(\mathcal{U}_2) = 5.3333$ and $\kappa(\mathcal{S}_{\pi/2}) = 5.5371$. A visual inspection of an estimate of the sectional density suggests that the arrival directions are uniformly distributed

on the hemisphere. However, the p-value of the kurtosis test of modality is 1.21×10^{-4}, rejecting the null hypothesis of unimodality. Note the large sample size. Although the kurtosis of arrival directions is statistically significantly different from that of uniform distributions, it is not significantly different in the usual sense.

6. Discussion

We conclude with several remarks on multivariate kurtosis and future direction of investigation.

It has been known that Mardia's definition of multivariate kurtosis [25] has some limitations. The main criticism of Mardia's original multivariate kurtosis κ is that it does not take into account all of fourth moments $E(X_iX_jX_kX_l)$ $(i, j, k, l \in \{1, \ldots, d\})$ of the (standardized) random vector $\mathbf{X} = (X_1, \ldots, X_d)$ [21]. This results in several proposals of scalar-valued multivariate kurtosis [10,20,21,24] and matrix-valued multivariate kurtosis [18,23,28]. We point out that once the rotational symmetry is assumed, all of the information in the fourth moments is contained in $\|\mathbf{X}\|$. With the change of variable $\mathbf{X} = \|\mathbf{X}\|\mathbf{U}$, $\mathrm{E}(\mathbf{X}\mathbf{X}^\top \otimes \mathbf{X}\mathbf{X}^\top) = \mathrm{E}\|\mathbf{X}\|^4\mathrm{E}(\mathbf{U}\mathbf{U}^\top \otimes \mathbf{U}\mathbf{U}^\top)$, where $\mathbf{U} = \mathbf{X}/\|\mathbf{X}\|$ follows the uniform distribution on $\mathbb{S}^{d-1}$, and $\otimes$ refers to the Kronecker product. Therefore, these various measures of multivariate kurtosis only depend on the fourth moment of $\|\mathbf{X}\|$ for rotationally symmetric distributions.

For the rotationally symmetric cases, the simple form of Mardia's kurtosis is clearly advantageous. As a comparison, Koziol's alternative definition of multivariate kurtosis [21] is $\kappa_K = \mathrm{E}(\mathbf{X}^\top\mathbf{Y})^4$, for independent $\mathbf{X}, \mathbf{Y} \sim f$. With the change of variable used above, it can be shown that $\kappa_K = 3\kappa^2/\{d(d+2)\}$. That is, both κ and κ_K measure the essentially same information. However, the empirical version of κ_K, $\hat{\kappa}_K = n^{-2}\sum_{i=1}^n\sum_{j=1}^n(\mathbf{X}_i^\top\mathbf{X}_j)^4$, is unappealing. Due to the double summation, the computation times for $\hat{\kappa}_K$ is $O(n^2d)$, compared with $O(nd)$ for Mardia's. Furthermore, $\hat{\kappa}_K$ is a biased estimator of κ with bias $n^{-1}(\mathrm{E}\|\mathbf{X}\|^8 - \kappa_K)$. Even for a large sample size and with a bounded support, the bias seems noneligible as the eight moment is typically large when the dimension is high.

We have limited ourselves with a perhaps strong assumption of rotational symmetry. The assumption greatly simplifies the technical arguments and serves well in our application to dimension reduction of directional data. We conjecture that similar results hold for both Euclidean and spherical elliptical distributions. Such an investigation will be the first step in developing moment-based exploration of directional data.

Matrix-valued multivariate kurtosis and multivariate skewness have been used in exploration of distributional features, such as tail behavior, presence of outliers, and departure from normality. To date, there has been no attempt of devising a matrix-valued spherical kurtosis or skewness for directional data. Using the tangent coordinates is an obvious starting point for this line of research.

7. Technical details

Proof of Theorem 1. Let $a_p = a_p(g) = (p+1)c_g(p)$, for $p \in \{0, 1, \ldots\}$.

Part (i). By Lemma 2, (5) holds if and only if

$$\frac{c_g(d+3)c_g(d-1)}{[c_g(d+1)]^2} \geq \frac{(d+2)^2}{d(d+4)},$$

which is equivalent to

$$a_{d+3}a_{d-1} \geq a_{d+2}^2. \tag{13}$$

We say that a sequence $\alpha = (a_p)_{p\geq 0}$ is logarithmically convex (log-convex, for short) if $a_pa_{p+2} \geq a_p^2$, and is strictly log-convex if $a_pa_{p+2} > a_p^2$, for all $p \geq 0$. If α is log-convex, then (13) holds for all $d \geq 0$. If α is strictly log-convex, then the inequality is strict as well. A logarithmically concave, or log-concave, sequence is defined similarly.

Assume without loss of generality that g is right-continuous. Define a nonnegative measure G on $(0, \infty)$ by $G((0, x]) = g(0) - g(x)$ for all $x > 0$. Then $G((x, \infty)) = g(x)$. With a slight abuse of notation, write $G(x) = G((0, x])$, then G is non-negative, non-decreasing and right-continuous. G is degenerate only if g is a positive constant on $[0, c)$ for some $c > 0$ and 0 on $[c, \infty)$.

Case I: Suppose that G is non-degenerate. Observe that for any $p \geq 0$,

$$\begin{aligned} a_p(g) &= \int_0^\infty (p+1)r^p g(r)dr = \int_0^\infty (p+1)r^p \int_0^\infty 1_{r\leq y}dG(y)dr \\ &= \int_0^\infty \int_0^y (p+1)r^p drdG(y) = \int_0^\infty y^{p+1}dG(y). \end{aligned}$$

Since $\|G\|_{\max} = g(0)$ is finite, $F = G/\|G\|_{\max}$ is a non-degenerate probability distribution function on $[0, \infty)$. Let $Y \sim F$, then the sequence $\beta = (b_k)_{k\geq 0}$, where $b_p = a_{p-1}(g)/\|G\|_{\max} = \mathrm{E}(Y^p)$, is a Stieltjes moment sequence [36]. Let $H(\beta)$ be the Hankel matrix generated by β, i.e., the (i, j)th element of $H(\beta)$ is b_{i+j}. It is known that a Stieltjes moment sequence has a totally positive Hankel matrix (see, e.g., Theorem 4.4 of [31]). In particular, all minors of order 2 of $H(\beta)$ is positive. That is, $b_pb_{p+2} > b_{p+1}^2$, for all $p \geq 0$, which in turn leads that the sequence $a_p(g)$ is also strictly log-convex.

Case II: Suppose that G is degenerate, which occurs only when f_g is a ball uniform, or $g(r) = c_0 1_{r<c_1}$ for some $c_0, c_1 > 0$. For such g, the sequence $a_p(g)$ is linear, and the equality of (13) holds. This completes a proof for Part (i).

Part (ii). Using a similar argument used above, (6) holds if $a_{d+3}a_{d-1} \le a_{d+2}^2$, which is implied by $a_{p+2}a_p \le a_{p+1}^2$ for any $p \in \{0, 2, \ldots\}$. It only remains to show that under the conditions on g, the sequence a_p is log concave.

Let $m = \sup_r g(r) = g(c)$. Without loss of generality, assume that g is right-continuous and $c = 1$. Then since the function $H : [0, 1] \to [0, 1]$, $H(\cdot) = m^{-1}g(\cdot)$, is a probability distribution function on $[0, 1]$,

$$a_p = m(p+1)\int_0^1 r^p m^{-1} g(r)dr = m\left\{1 - \int_0^1 y^p dH(y)\right\}.$$

Notice that the sequence $(c_p)_{p\ge 0}$, $c_p = \int_0^1 y^p dH(y)$ is a Hausdorff moment sequence. Hausdorff showed that (c_p) is a Hausdorff moment sequence if and only if it is completely monotone [Theorem 1.5, 36]. A sequence (c_p) is said to be completely monotone if for all $j \ge 0$, $p \ge 0$,

$$\Delta^j c_p := \sum_{i=1}^{j} (-1)^i \binom{j}{i} a_{p+i} \ge 0.$$

In particular, by plugging in $j = 1, 2$, we have $c_p \ge c_{p+1}$, and $c_p + c_{p+2} \ge 2c_{p+1}$. Together with the fact that $c_0 = 1$, we conclude that (c_p) is convex and $c_p \le 1$ for any p. This immediately implies that a_p is nonnegative and concave. An application of Jensen's inequality leads that $\log a_p$ is concave as well. □

Proof of Lemma 3. We utilize the hyperspherical coordinate system for $\mathbf{x} \in \mathbb{S}^d$, with respect to $\boldsymbol{\mu}$. Without loss of generality, assume that $\boldsymbol{\mu} = \mathbf{e}_1$. For any $\mathbf{x} = (x_1, \ldots, x_{d+1})^\top \in \mathbb{S}^d$, we write $x_1 = \cos\phi_1$, $x_k = \left(\prod_{i=1}^{k-1} \sin\phi_i\right)\cos\phi_k$, $k = 2, \ldots, d$, and $x_{d+1} = \prod_{i=1}^{d} \sin\phi_i$, for $\phi_k \in [0, \pi]$, $k = 1, \ldots, d-1$, and $\phi_d \in [0, 2\pi)$. The volume element of $\mathbb{S}^d$ is then $dV_d = \prod_{i=1}^{d-1}(\sin\phi_i)^{d-i}d\boldsymbol{\phi}$. So, the density function of $\mathbf{X}$ with respect to $\boldsymbol{\phi}$ is

$$f_{\boldsymbol{\phi}}(\boldsymbol{\phi}) = g_S(\phi_1)\prod_{i=1}^{d-1}(\sin\phi_i)^{d-i}. \tag{14}$$

The log-mapped $\mathbf{X}$ is then

$$\mathbf{Y} = \mathrm{Log}_{\boldsymbol{\mu}}(\mathbf{X}) = \frac{\phi_1}{\sin\phi_1}(x_2, \ldots, x_{d+1})^\top = \phi_1 \begin{pmatrix} \cos\phi_2 \\ \sin\phi_2\cos\phi_3 \\ \vdots \\ \left(\prod_{i=2}^{d-1}\sin\phi_i\right)\cos\phi_d \\ \left(\prod_{i=2}^{d-1}\sin\phi_i\right)\sin\phi_d \end{pmatrix},$$

and the Jacobian of the transformation $\boldsymbol{\phi} \mapsto \mathrm{Log}_{\boldsymbol{\mu}}(\mathbf{X})$ is $\mathbf{J} = \phi_1^{d-1}\prod_{i=2}^{d-1}(\sin\phi_i)^{d-i}$. Observe that $\|\mathbf{Y}\| = \phi_1$. Thus, the density function of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\boldsymbol{\phi}}(\boldsymbol{\phi})\mathbf{J}^{-1} = g_S(\phi_1)\frac{(\sin\phi_1)^{d-1}}{\phi_1^{d-1}} = g_S(\|\mathbf{y}\|)\left(\frac{\sin\|\mathbf{y}\|}{\|\mathbf{y}\|}\right)^{d-1}, \tag{15}$$

which proves (i).

A proof for (ii) is derived from observing that $\|\mathbf{Y}\| = \phi_1$ is independent of $\phi_2, \ldots, \phi_d$ as shown in (14).

For (iii), We observe from (15) that $f_{\mathbf{Y}}$ on $\mathbb{R}^d$ is rotationally symmetric about the origin and that the sectional density of $f_{\mathbf{Y}}$ along any ray from the origin is

$$g(r) = g_S(r)\left(\frac{\sin r}{r}\right)^{d-1}.$$

We then follow the proof of Lemma 2. In particular, by Lemma 1, $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_d$ for $\sigma^2 = d^{-1}\mathrm{E}\|\mathbf{Y}\|^2$. Then, $\kappa(f_{\mathbf{X}}) = \kappa(f_{\mathbf{Y}}) = \mathrm{E}(\mathbf{Y}^\top\boldsymbol{\Sigma}^{-1}\mathbf{Y})^2 = d^2\mathrm{E}\|\mathbf{Y}\|^4/(\mathrm{E}\|\mathbf{Y}\|^2)^2$. Writing $\mathrm{E}\|\mathbf{Y}\|^m = C(m, d; g_S)/C(0, d; g_S)$ and rearranging gives (iii). □

Proof of Lemma 4. For a given $\epsilon > 0$, suppose that

$$\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) < \epsilon. \tag{16}$$

To avoid the non-converging $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{x}\|$ case, let us assume for now that $\mathbf{x} \in \mathbb{S}^d$ satisfy

$$\rho(-\boldsymbol{\mu}, \mathbf{x}) > 2\epsilon. \tag{17}$$

Write $\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x} = \rho(\boldsymbol{\mu},\mathbf{x})\mathbf{v}(\boldsymbol{\mu})$, where $\mathbf{v}(\boldsymbol{\mu}) = \{\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\mu}^\top\mathbf{x})\}/\|\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\mu}^\top\mathbf{x})\|$ is the spherical projection of $\boldsymbol{\mu}$ onto the nullspace of $\mathbf{x}$. Denote the geodesic segment from $\mathbf{x}$ to $-\mathbf{x}$ passing through $\boldsymbol{\mu}$ by $\gamma(\boldsymbol{\mu})$. It can be checked that $\gamma(\boldsymbol{\mu})$ also passes $\mathbf{v}(\boldsymbol{\mu})$ and that the angle formed by the two geodesic segments $\gamma(\boldsymbol{\mu})$ and $\gamma(\hat{\boldsymbol{\mu}})$ (either at $\mathbf{x}$ or at $-\mathbf{x}$) equals $\rho\{\mathbf{v}(\boldsymbol{\mu}), \mathbf{v}(\hat{\boldsymbol{\mu}})\}$. Under the conditions (16) and (17), we will construct an upper bound for the angle.

Specifically, the spherical law of cosines for the spherical triangle formed by the vertices $\boldsymbol{\mu}$, $\hat{\boldsymbol{\mu}}$ and $-\mathbf{x}$ states that $\cos C = (\cos c - \cos a \cos b)/(\sin a \sin b)$, where $C = \rho\{\mathbf{v}(\boldsymbol{\mu}), \mathbf{v}(\hat{\boldsymbol{\mu}})\}$, $c = \rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) < \epsilon$, $a = \rho(\boldsymbol{\mu}, -\mathbf{x}) > 2\epsilon$ and $b = \rho(\hat{\boldsymbol{\mu}}, -\mathbf{x})$. Since $a > c$, $b \in [a - c, a + c]$ by the triangle inequality. To find an upper bound of C, or equivalently a lower bound of $\cos C$, we choose b so that $\cos C$ is minimized. Taking the derivative of $\cos C$ with respect to b, we have for $h(b) := \cos C$, $h'(b) = -(\sin b)^{-2}(\cos b \cos c - \cos a)$. The solution of $h'(b) = 0$ on $[a - c, a + c]$ is unique and satisfies

$$\cos b = \frac{\cos a}{\cos c}.$$

We note that the above choice of b, for given values of a and c, makes a right-angled triangle. In particular, the angle formed at the vertex $\boldsymbol{\mu}$ is $\pi/2$.

All in all, we have

$$\cos C \geq \frac{\cos c - \cos a \frac{\cos a}{\cos c}}{\sin a \left\{1 - (\frac{\cos a}{\cos c})^2\right\}^{1/2}} = \left(1 + \frac{\cos^2 c - 1}{\sin^2 a}\right)^{\frac{1}{2}},$$

and since $\cos c > \cos\epsilon$ and $\sin^2 a \leq 1$, $\cos C > \cos\epsilon$. We conclude that the conditions (16) and (17) imply that $C = \rho\{\mathbf{v}(\boldsymbol{\mu}), \mathbf{v}(\hat{\boldsymbol{\mu}})\} < \epsilon$ and that $\|\mathbf{v}(\boldsymbol{\mu}) - \mathbf{v}(\hat{\boldsymbol{\mu}})\| = \sqrt{2 - 2\cos C} < \sqrt{2 - 2\cos\epsilon} = 2\sin(\epsilon/2)$. Then

$$\begin{aligned}\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{x}\| &= \|\rho(\boldsymbol{\mu},\mathbf{x})\mathbf{v}(\boldsymbol{\mu}) - \rho(\hat{\boldsymbol{\mu}},\mathbf{x})\mathbf{v}(\hat{\boldsymbol{\mu}})\| \\ &= \|\rho(\boldsymbol{\mu},\mathbf{x})\mathbf{v}(\boldsymbol{\mu}) - \rho(\hat{\boldsymbol{\mu}},\mathbf{x})\mathbf{v}(\boldsymbol{\mu}) + \rho(\hat{\boldsymbol{\mu}},\mathbf{x})\mathbf{v}(\boldsymbol{\mu}) - \rho(\hat{\boldsymbol{\mu}},\mathbf{x})\mathbf{v}(\hat{\boldsymbol{\mu}})\| \\ &\leq |\rho(\boldsymbol{\mu},\mathbf{x}) - \rho(\hat{\boldsymbol{\mu}},\mathbf{x})|\|\mathbf{v}(\boldsymbol{\mu})\| + |\rho(\hat{\boldsymbol{\mu}},\mathbf{x})|\|\mathbf{v}(\boldsymbol{\mu}) - \mathbf{v}(\hat{\boldsymbol{\mu}})\| < \rho(\boldsymbol{\mu},\hat{\boldsymbol{\mu}}) + 2\pi\sin(\epsilon/2) \leq c_0\epsilon, \end{aligned} \tag{18}$$

for a constant $c_0 < \infty$.

By definition, $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}\| = O(n^{-1/2})$ if for any $\epsilon > 0$, there are $M, N < \infty$ such that for every $n \geq N$, $\Pr(\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}\| < M/\sqrt{n}) > 1 - \epsilon$. Fix ϵ. Since $\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = O_p(n^{-1/2})$, choose M, N_0 such that $\Pr(\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) < M/\sqrt{n}) \geq 1 - \epsilon/2$ for $n \geq N_0$. Given the events that $\rho(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) < M/\sqrt{n}$ and $\rho(-\boldsymbol{\mu}, \mathbf{X}) > 2M/\sqrt{n}$, hereafter denoted as $E_{1(n)}$ and $E_{2(n)}$, (18) leads that $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{x}\| \leq c_0/\sqrt{n}$, for some $c_0 < \infty$. Therefore,

$$\begin{aligned}\Pr(\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{x}\| \leq c_0/\sqrt{n}) &\geq \Pr(\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{x}\| \leq c_0/\sqrt{n} \mid E_{1(n)} \cap E_{2(n)}) \Pr(E_{1(n)} \cap E_{2(n)}) \\ &= 1 \cdot \Pr(E_{1(n)} \cap E_{2(n)}) \geq \Pr(E_{1(n)}) + \Pr(E_{2(n)}) - 1 \geq -\epsilon/2 + \Pr(E_{2(n)}). \end{aligned} \tag{19}$$

Since $\rho(\boldsymbol{\mu}, \mathbf{X})$ has an absolutely continuous distribution on $(0, \pi)$, one can choose $N \geq N_0$ large enough so that $\Pr(E_{2(n)}) = \Pr(\rho(-\boldsymbol{\mu}, \mathbf{X}) > 2M/\sqrt{n}) = \Pr(\rho(\boldsymbol{\mu}, \mathbf{X}) \leq \pi - 2M/\sqrt{n}) > 1 - \epsilon/2$. This, with (19), shows that $\|\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{X} - \mathrm{Log}_{\hat{\boldsymbol{\mu}}}\mathbf{X}\| = O(n^{-1/2})$. □

CRediT authorship contribution statement

Byungwon Kim: Conceptualization, Methodology, Software, Writing - original draft. **Jörn Schulz:** Formal analysis, Data curation. **Sungkyu Jung:** Supervision, Validation, Investigation, Visualization, Project administration, Writing - review & editing, Funding acquisition.

Acknowledgments

Jung's research was supported by Research Resettlement Fund for the new faculty of Seoul National University, South Korea and the NRF, Korea grant funded by the Korea government (MSIT) (No. 2019R1A2C2002256).

References

[1] R. Abbasi, Y. Abdou, T. Abu-Zayyad, J. Adams, J.A. Aguilar, M. Ahlers, D. Altmann, K. Andeen, J. Auffenberg, X. Bai, et al., IceCube Collaboration, Search for a diffuse flux of astrophysical muon neutrinos with the IceCube 40-string detector, Phys. Rev. D 84 (8) (2011) 082001.
[2] G. Alsmeyer, Nonnegativity of odd functional moments of positive random variables with decreasing density, Statist. Probab. Lett. 26 (1) (1996) 75–82.
[3] R. Bhattacharya, L. Lin, Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces, Proc. Amer. Math. Soc. 145 (1) (2017) 413–428.
[4] C. Bingham, K.V. Mardia, A small circle distribution on the sphere, Biometrika 65 (2) (1978) 379–389.
[5] J. Damon, J. Marron, Backwards principal component analysis and principal nested relations, J. Math. Imaging Vision 50 (1–2) (2014) 107–114.
[6] I.L. Dryden, K.-R. Kim, C.A. Laughton, H. Le, Principal nested shape space analysis of molecular dynamics data, Ann. Appl. Stat. 13 (4) (2019) 2213–2234.
[7] I.L. Dryden, K.V. Mardia, Statistical Shape Analysis, second ed., Wiley, Chichester, 2016.

[8] B. Eltzner, S. Huckemann, K.V. Mardia, Torus principal component analysis with applications to RNA structure, Ann. Appl. Stat. 12 (2) (2018) 1332–1359.
[9] B. Eltzner, S. Jung, S. Huckemann, Dimension reduction on polyspheres with application to skeletal representations, in: International Conference on Geometric Science of Information, Springer, 2015, pp. 22–29,
[10] C. Franceschini, N. Loperfido, Some inequalities between measures of multivariate kurtosis, with application to financial returns, in: Mathematical and Statistical Methods for Actuarial Sciences and Finance, 2012, pp. 211–218,
[11] E. García-Portugués, D. Paindaveine, T. Verdebout, On optimal tests for rotational symmetry against new classes of hyperspherical distributions, J. Amer. Stat. Assoc. (2020) in press.
[12] D.K. Hildebrand, Kurtosis measures bimodality? Amer. Statist. 25 (1) (1971) 42–43.
[13] J. Hong, J. Vicory, J. Schulz, M. Styner, J.S. Marron, S.M. Pizer, Non-Euclidean classification of medically imaged objects via s-reps, Med. Image Anal. 18 (31) (2016) 37–45.
[14] S.F. Huckemann, B. Eltzner, Backward nested descriptors asymptotics with inference on stem cell differentiation, Ann. Statist. 46 (5) (2018) 1994–2019.
[15] S. Jung, I.L. Dryden, J.S. Marron, Analysis of principal nested spheres, Biometrika 99 (3) (2012) 551–568.
[16] S. Jung, M. Foskey, J.S. Marron, Principal arc analysis on direct product manifolds, Ann. Appl. Stat. 5 (1) (2011) 578–603.
[17] B. Kim, S. Huckemann, J. Schulz, S. Jung, Small-sphere distributions for directional data with application to medical imaging, Scand. J. Stat. 46 (2019) 1047–1071.
[18] T. Kollo, Multivariate skewness and kurtosis measures with an application in ICA, J. Multivariate Anal. 99 (10) (2008) 2328–2338.
[19] T. Kollo, A. Roos, On kotz-type elliptical distributions, in: Contemporary Multivariate Analysis and Design of Experiments, World Scientific, 2005, pp. 159–170.
[20] J.A. Koziol, An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives, Metrika 34 (1987) 17–24.
[21] J.A. Koziol, A note on measures of multivariate kurtosis, Biom. J. 31 (1989) 619–624.
[22] C. Ley, Y. Swan, B. Thiam, T. Verdebout, Optimal R-estimation of a spherical location, Statist. Sinica 23 (1) (2013) 305–332.
[23] N. Loperfido, A new kurtosis matrix, with statistical applications, Linear Algebra Appl. 512 (2017) 1–17.
[24] N. Loperfido, Some remarks on Koziol's kurtosis, J. Multivariate Anal. 175 (104565) (2020).
[25] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (3) (1970) 519–530.
[26] K.V. Mardia, P.E. Jupp, Directional Statistics, Wiley, New York, 2000, p. 429.
[27] R.K. McClure, M. Styner, E. Maltbie, J.A. Liebermann, S. Gouttard, G. Gerig, X. Shi, H. Zhu, Localized differences in caudate and hippocampal shape are associated with schizophrenia but not antipsychotic type, Psychiatry Research: Neuroimaging 211 (1) (2013) 1–10.
[28] T.F. Mori, V.K. Rohatgi, G.J. Szekely, On multivariate skewness and kurtosis, Theory Probab. Appl. 38 (1993) 547–551.
[29] D. Paindaveine, T. Verdebout, Optimal rank-based tests for the location parameter of a rotationally symmetric distribution on the hypersphere, in: Mathematical Statistics and Limit Theorems, Springer, 2015, pp. 249–269,
[30] D. Paindaveine, T. Verdebout, et al., Inference on the mode of weak directional signals: a Le Cam perspective on hypothesis testing near singularities, Ann. Statist. 45 (2) (2017) 800–832.
[31] A. Pinkus, Totally Positive Matrices, Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2009.
[32] S.M. Pizer, S. Jung, D. Goswami, X. Zhao, R. Chaudhuri, J.N. Damon, S. Huckemann, J.S. Marron, Nested sphere statistics of skeletal models, in: M. Breus, A. Bruckstein, P. Maragos (Eds.), Proc. Dagstuhl Workshop on Innovations for Shape Analysis: Models and Algorithms, in: Innovations for Shape Analysis: Models and Algorithms, Springer, New York, 2013, pp. 93–115.
[33] U. Rösler, Distributions slanted to the right, Stat. Neerl. 49 (1) (1995) 83–93.
[34] J. Schulz, S.M. Pizer, J.S. Marron, F. Godtliebsen, Non-linear hypothesis testing of geometric object properties of shapes applied to hippocampi, J. Math. Imaging Vision 54 (2016) 15–34.
[35] X. Shi, J.G. Ibrahim, J. Lieberman, M. Styner, Y. Li, H. Zhu, Two-stage empirical likelihood for longitudinal neuroimaging data, Ann. Appl. Stat. 5 (2B) (2011) 1132–1158.
[36] J.A. Shohat, J.D. Tamarkin, The Problem of Moments, American Mathematical Scociety, Providence, 1943.
[37] P.H. Westfall, Kurtosis as peakedness, 1905–2014. R.I. P., Amer. Statist. 68 (3) (2014) 191–195.
[38] Q. Yu, X. Lu, J. Marron, Principal nested spheres for time-warped functional data analysis, J. Comput. Graph. Statist. 26 (1) (2017) 144–151.