# Small sphere distributions for directional data with application to medical imaging

Byungwon Kim*[1], Stephan Huckemann†[2], Jörn Schulz‡[3] and Sungkyu Jung§[1]

[1]Department of Statistics, University of Pittsburgh
[2]Institute for Mathematical Stochastics, University of Göttingen
[3]Department of Electrical and Computer Engineering, University of Stavanger

March 3, 2017

**Abstract**

We propose new distribution families, called small-sphere distributions, for modeling multivariate directional data on $(\mathbb{S}^{p-1})^K$ for $p \geq 3$ and $K \geq 1$. In a special case of univariate directions in $\Re^3$, the new densities model random directions on $\mathbb{S}^2$ with a tendency to vary along a small circle on the sphere, and with a unique mode on the small circle. The proposed multivariate densities are capable of modeling multivariate directions which are dependent to each other. The proposed distributions are useful in, for example, modeling object shapes in medical imaging, where multivariate directions are used to represent the shape of a 3-dimensional objects. The directional data tend to follow the proposed small-sphere distributions when the underlying objects are randomly and rotationally deformed. The proposed models have several advantages over other methods of analyzing small-circle-concentrated data, including formal procedures to make inferences on the association and small-circle fittings. We demonstrate the use of the multivariate small-sphere distributions in analysis of skeletally-represented object shapes and human knee gait data.

**Keywords**: Bingham-Mardia distribution, directional data, skeletal representation, small circle, small sphere, likelihood ratio test, maximum likelihood estimation, von Mises-Fisher distribution.

*byk4@pitt.edu
†huckeman@math.uni-goettingen.de
‡jorn.schulz@uit.no
§sungkyu@pitt.edu

# 1 Introduction

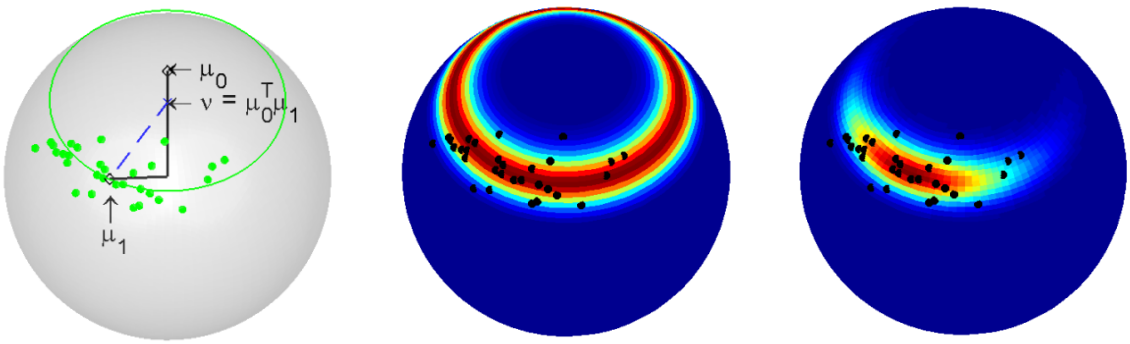
In medical imaging, accurately assessing and correctly diagnosing shape changes of internal organs is a major objective of a substantial challenge. Shape deformations can occur through long-term growth or necrosis as well as by short-term natural deformations. In view of surgery and radiation therapy, it is important to model all possible variations of deformations of an object by both long- and short-term changes in order to control its exact status and shape at treatment time. *Rotational deformations* such as rotation, bending, and twisting form a key subcategory of possible shape changes. For instance, shape changes of hippocampi in the human brain have been shown to mainly occur in the way of bending and twisting (Joshi et al., 2002; Pizer et al., 2013).

For the task of modeling 3D objects, an abundance of approaches has been introduced. Closely related to our work are landmark-based shape models (Cootes et al., 1992; Dryden and Mardia, 1998; Kurtek et al., 2011) where a solid object is modeled by the positions of surface points, chosen either anatomically, mathematically or randomly. A richer family of models is obtained by attaching directions normal to the sampled surface points. More generally, in skeletal representations (called *s-reps*, Siddiqi and Pizer, 2008), an object is modeled by the combination of skeletal positions (lying on a medial sheet inside of the object) and spoke vectors (connecting the skeletal positions with the boundary of the object). In these models, describing the variation of rotational deformations can be transformed into a problem of exploring the motion of *directional vectors* on the unit two-sphere. As argued in Schulz et al. (2015), directional vectors representing rotational deformations tend to be concentrated on small circles on the unit sphere; a toy data example in Fig. 1 shows a typical pattern of such observations.

Motivated by the analysis of such s-rep data, we propose new distribution families and their multivariate extensions in order to model directional data on the unit sphere $\mathbb{S}^{p-1} = \{x \in \mathbb{R}^p \,|\, \|x\| = 1\}$ in arbitrary dimension $p \geq 3$. (Throughout the paper, $\|x\| = x^\top x$ is the usual 2-norm of the vector $x$.) To precisely describe the intended data situation, we define a $(p-2)$-dimensional *subsphere* of $\mathbb{S}^{p-1}$ by the set of all points equidistant from $\mu \in \mathbb{S}^{p-1}$, denote

$$\mathcal{C}(\mu, \nu) = \{x \in \mathbb{S}^{p-1} \,|\, \delta(\mu, x) = \cos^{-1}(\nu)\}, \quad \nu \in (-1, 1).$$

Here, $\delta(u, v) = \cos^{-1}(u^\top v)$ is the geodesic distance between $u, v \in \mathbb{S}^{p-1}$. The subsphere is called a *great subsphere* if $\nu = 0$ and a proper *small subsphere* if $\nu \neq 0$. Note that $\mathcal{C}(\mu, \nu) \subset \mathbb{S}^{p-1}$ is well-defined for all $p > 1$. For the special case of $p = 3$, $\mathcal{C}(\mu, \nu)$ is a circle, a *great circle* if $\nu = 0$, and a proper *small circle* if $\nu \neq 0$. To model the data in Fig. 1, one may naively use the Bingham-Mardia distribution (BM, Bingham and Mardia, 1978), which is a family of densities on $\mathbb{S}^2$ with a modal ridge along a small circle. However, typical observations we encountered in




(a) Toy data (b) Bingham-Mardia (c) Proposed density

Figure 1: A toy example showing observations on the unit sphere, $\mathbb{S}^2$. (a) Directional vectors, displayed on $\mathbb{S}^2$ as solid green points, are distributed near a small circle. The heat maps of fitted Bingham-Mardia density and the proposed small-sphere density (of the first kind) are overlaid. Red: high density, blue: low density.

applications do not spread over the circle, and the BM distribution does not fit well, as shown in Fig. 1(b). Moreover, when multiple direction vectors are available from a single observation, that is, data are on a polysphere $(\mathbb{S}^2)^K$, there is no available tool to model dependence between directions.

In this paper, we propose two types of new distribution families for random direction vectors on $\mathbb{S}^{p-1}$, which we call *small-sphere distributions* of the first (S1) and second (S2) kind. In the usual case $p = 3$, the proposed distributions may be called *small-circle distributions*. These two distribution families are designed to have higher densities on $\mathcal{C}(\mu, \nu)$ and have a unique mode on $\mathcal{C}(\mu, \nu)$. An example of a small-sphere density, fitted to the toy data is shown in Fig. 1(c). The new densities are natural extensions of the BM distribution with an additional term explaining a decay from a mode. If the additional term is a von Mises-Fisher (vMF) density on $\mathbb{S}^{p-1}$, we get the S1, which is a subfamily of the general Fisher-Bingham distribution (Mardia and Jupp, 2000). On the other hand, if the additional term is a vMF density on the small sphere ($\cong \mathbb{S}^{p-2}$), we get the S2 distribution, in which case the "horizontal" and "vertical" components of the direction vectors are independent of each other.

Several multivariate extensions of the new distributions to $(\mathbb{S}^{p-1})^K$, $K \geq 2$, are discussed as well. In particular, we show that a special case, called *MS2*, of our multivariate extensions is capable of modeling *dependent* random vectors, and has straightforward interpretation and fast estimation of its parameters. This MS2 distribution is specifically designed with s-rep applications in mind. In particular, s-rep data from rotationally-deformed objects have directional vectors that are "rotated together," share a common axis of rotation, and are "horizontally dependent" (when the axis is considered to be vertically positioned). The component-wise inde-

| | First kind (S1) | Second kind (S2) | |
|---|---|---|---|
| Univariate | S1 | S2 | |
| Multivariate (indep.) | iMS1 | iMS2 | |
| Multivariate (dep.) | x | GMS2 | MS2 |
| Simulation | Gibbs sampling | x | Exact sampling |
| Estimation | Approximate m.l.e. | x | Approximate m.l.e. |
| Hypothesis testing | Likelihood ratio | x | Likelihood ratio |

Table 1: Small-sphere distributions developed in this paper. Items with "x" marks are not addressed.

pendence of the S2 distributions plays a key role in this simple and interpretable extension. The likelihood-based parameter estimation of the multivariate distributions and testing procedures are also discussed.

The contribution of this paper is summarized in Table 1. While the new distributions contribute to the literature of directional distributions (Mardia and Jupp, 2000), the proposed estimation procedures for the S1, S2 and MS2 parameters can be thought of as a method of fitting small-subspheres to data, which has been of separate interest. Nonparametric least-squares type solutions for such problem dates back to Mardia and Gadsden (1977), Gray et al. (1980), and Rivest (1999). Jung et al. (2012) proposed to recursively fitting small-subspheres in dimension reduction of directional and shape data. Pizer et al. (2013) proposed to combine separate small-circle fitting results in the analysis of s-rep data. In a similar spirit, Jung et al. (2011) and Schulz et al. (2015) also considered fitting small-circles in applications to s-rep analysis. Our estimators perform better, with smaller mean angular errors in small-circle fit, than the recent developments. Moreover, our parametric framework provides a formal large-sample likelihood ratio test procedure, which is applicable to a number of hypothesis settings and is more powerful in testing the dependence among multivariate directions than natural alternatives.

The rest of the article is organized as follows. In Section 2, we introduce the proposed densities of the S1 and S2 distributions and discuss their multivariate extensions including the MS2 distribution. Procedures of obtaining random observations from the proposed distributions are also discussed. In Section 3, algorithms to obtain approximate maximum likelihood estimators of the parameters are discussed. In Section 4, we introduce several hypotheses of interest and procedures of large-sample approximated likelihood-ratio tests. In Section 5, we empirically show that the proposed model is superior at estimating the dependence in the multivariate model and at the small-circle estimation, and show that the proposed testing procedure is effective at preventing an overfitting. In Sections 6 and 7 we demonstrate applications of the new multivariate distributions to analysis of models that represent human organs and knee motions. Appendix and the online supplementary material contain technical details and additional

numerical results.

## 2 Parametric small-sphere models

First we introduce two classical spherical densities, then we suitably combine them for our purposes.

### 2.1 Two classical distributions on $\mathbb{S}^{p-1}$

The von Mises-Fisher (vMF) distribution (Mardia and Jupp, 2000, p.168) is a fundamental unimodal and isotropic distribution for directions with density

$$f_{\mathrm{vMF}}(x;\mu,\kappa) = \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{\Gamma(p/2) I_{p/2-1}(\kappa)} \exp\{\kappa \mu^\top x\}, \quad x \in \mathbb{S}^{p-1}\,. \tag{1}$$

Here, $\Gamma$ is the gamma function and $I_v$ is the modified Bessel function of the first kind and order $v$. The parameter $\mu \in \mathbb{S}^{p-1}$ locates the unique mode with $\kappa \geq 0$ representing the degree of concentration.

The Bingham-Mardia (BM) distribution was introduced by Bingham and Mardia (1978) to fit data in $\mathbb{S}^2$ that cluster near a small circle $\mathcal{C}(\mu,\nu)$. For an arbitrary dimension $p \geq 3$, the BM density is given by

$$f_{\mathrm{BM}}(x;\mu,\kappa,\nu) = \frac{1}{\alpha(\kappa,\nu)} \exp\{-\kappa(\mu^\top x - \nu)^2\}, \quad x \in \mathbb{S}^{p-1}, \tag{2}$$

where $\alpha(\kappa,\nu) > 0$ is the normalizing constant.

For our purpose of generalizing these distributions, we represent the variable $x \in \mathbb{S}^{p-1}$, $p \geq 3$, by spherical angles $\phi_1, \ldots, \phi_{p-1}$ satisfying $\cos\phi_1 = \mu^\top x$. Setting $s := \cos\phi_1 \in [-1,1]$ and $\phi := (\phi_2, \ldots, \phi_{p-1}) \in [0,\pi]^{p-3} \times [0, 2\pi)$, the random vector $(s,\phi)$ following the von Mises-Fisher (1) or Bingham-Mardia (2) distribution has the respective density:

$$g_{\mathrm{vMF}}(s,\phi;\kappa) = \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{\Gamma(p/2) I_{p/2-1}(\kappa)} \exp\{\kappa s\}\,, \tag{3}$$

$$g_{\mathrm{BM}}(s,\phi;\kappa,\nu) = \frac{1}{\alpha(\kappa,\nu)} \exp\{-\kappa(s-\nu)^2\}\,. \tag{4}$$

In consequence, for both distributions, $s$ and $\phi$ are independent, and the marginal distribution of $\phi$, which parametrizes a co-dimension 1 unit sphere $\mathbb{S}^{p-2}$, is uniform. In (3), the marginal distribution of $s$ is a shifted exponential distribution truncated to $s \in [-1,1]$, while in (4) the marginal distribution of $s$ is a normal distribution truncated to $s \in [-1,1]$. Both densities are

isotropic, i.e. rotationally symmetric with respect to $\mu$. The vMF density is maximal at the mode $\mu$ and decreases as the latitude $\phi_1$ increases, while the BM density is uniformly maximal on the small-sphere $\mathcal{C}(\mu, \nu)$ and decreases as $\phi_1$ deviates from $\cos^{-1}(\nu)$.

### 2.2 Small-sphere distributions of the first and second kind

The proposed small-sphere densities of the first and second kind on $\mathbb{S}^{p-1}$, for $x = (x_1, \ldots, x_p) \in \mathbb{S}^{p-1}$ with parameters $\mu_0, \mu_1 \in \mathbb{S}^{p-1}$, $\nu = \mu_0^\top \mu_1 \in (-1, 1)$, $\kappa_0 > 0$, $\kappa_1 > 0$, are given by

$$f_{\mathrm{S1}}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \frac{1}{a(\kappa_0, \kappa_1, \nu)} \exp\{-\kappa_0(\mu_0^\top x - \nu)^2 + \kappa_1 \mu_1^\top x\}, \tag{5}$$

$$f_{\mathrm{S2}}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \frac{1}{b(\kappa_0, \kappa_1, \nu)} \exp\left\{-\kappa_0(\mu_0^\top x - \nu)^2 + \kappa_1 \frac{\mu_1^\top P_{\mu_0} x}{\sqrt{\mu_1^\top P_{\mu_0} \mu_1 x^\top P_{\mu_0} x}}\right\}, \tag{6}$$

respectively, where $a(\kappa_0, \kappa_1, \nu)$ and $b(\kappa_0, \kappa_1, \nu)$ are normalizing constants. Here, $P_{\mu_0}$ denotes the matrix of orthogonal projection to the orthogonal *complement* of $\mu_0$; $P_{\mu_0} = I_p - \mu_0 \mu_0^\top$, where $I_p$ is the identity matrix. (In (6), we use the convention $0/0 = 0$.)

These distributions are well-suited to model observations that are concentrated near the small sphere $\mathcal{C}(\mu_0, \nu)$ but are not rotationally symmetric. The first kind (5) is a natural combination of the vMF (1) and BM (2) distributions. The parameter $\mu_0$ represents the axis of the small sphere $\mathcal{C}(\mu_0, \nu)$, while $\mu_1$ gives the mode of the distribution, which, by the definition of $\nu$, is on the small sphere $\mathcal{C}(\mu_0, \nu)$. The three parameters $(\mu_0, \mu_1, \nu)$ on $\mathbb{S}^2$ are illustrated in Fig. 1(a). The parameter $\kappa_0$ controls the *vertical concentration* towards the small sphere (with an understanding that $\mu_0$ is arranged vertically). In (5), $\kappa_1$ controls the isotropic part of the concentration around the mode, forcing the density to decay from $\mu_1$ in all directions on $\mathbb{S}^{p-1}$.

The rationale for the second kind (6) is better understood with a change of variables. Let us assume for now $\mu_0 = (1, 0, \ldots, 0)^\top$. Then for any $x = (x_1, \ldots, x_p)^\top \in \mathbb{S}^{p-1}$, write $s := x_1 = \mu_0^\top x$. If the spherical coordinate system $(\phi_1, \ldots, \phi_{p-1})$ as defined for (4) is used, then $s = \cos\phi_1$. The "orthogonal complement" of $s$ is denoted by

$$y := (x_2, \ldots, x_p)/\sqrt{1 - s^2} \in \mathbb{S}^{p-2}, \tag{7}$$

where the vector $y$ is obtained from the relation $P_{\mu_0} x / \|P_{\mu_0} x\| = (0, y) \in \mathbb{S}^{p-1}$. Similarly, define $\widetilde{\mu_1} \in \mathbb{S}^{p-2}$ as the last $p-1$ coordinates of $P_{\mu_0}\mu_1/\|P_{\mu_0}\mu_1\|$. Then the random vector

$(s, y) \in [-1, 1] \times \mathbb{S}^{p-2}$ from the S1 or S2 has densities

$$g_{\mathrm{S1}}(s, y; \mu_1, \kappa_0, \kappa_1) = \frac{1}{a(\kappa_0, \kappa_1, \nu)} \exp\left\{-\kappa_0(s-\nu)^2 + \kappa_1 \mu_1^\top \left(s, \sqrt{1-s^2}y\right)\right\}, \quad (8)$$

$$g_{\mathrm{S2}}(s, y; \mu_1, \kappa_0, \kappa_1) = \frac{1}{b(\kappa_0, \kappa_1, \nu)} \exp\left\{-\kappa_0(s-\nu)^2 + \kappa_1 \widetilde{\mu_1}^\top y\right\}, \quad (9)$$

respectively, for $s \in [-1, 1]$, $y \in \mathbb{S}^{p-2}$. The subtle difference is that for the first kind (8), the "vMF part" (the second term in the exponent) is not statistically independent from the "BM part", while it is true for the second kind (9). That is, $s$ and $y$ are independent only in the second kind. Accordingly, in (9), $\kappa_1$ controls the *horizontal concentration* towards the mode $\mu_1$. The parameters $\mu_0, \mu_1$ and $\kappa_0$ of the second kind have the same interpretations as those of the first kind.

We use the notation $X \sim \mathrm{S1}(\mu_0, \mu_1, \kappa_0, \kappa_1)$ and $Y \sim \mathrm{S2}(\mu_0, \mu_1, \kappa_0, \kappa_1)$ for random directions $X, Y \in \mathbb{S}^{p-1}$ following small-sphere distributions of the first and second kind with parameters $(\mu_0, \mu_1, \kappa_0, \kappa_1)$, respectively. The proposed distributions are quite flexible and can fit a wide range of data. In Figure 2, we illustrate the S1 densities (5) with various values of the concentration parameters $\kappa_0, \kappa_1$. In all cases, the density is relatively high near the small circle $\mathcal{C}(\mu_0, \nu)$ and has the mode at $\mu_1 \in \mathcal{C}(\mu, \nu)$. Despite the difference in their formulations, the S2 densities (6) look similar to S1 densities for each fixed parameter-set. We refer to the online supplementary material for several visual examples of the S2 density.

The following invariance properties are proven in the Appendix.

**Proposition 1.** *Let $X, Y \in \mathbb{S}^{p-1}$ be a random direction with $X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$ and $Y \sim S2(\mu_0, \mu_1, \kappa_0, \kappa_1)$) and let $B$ be a full rank $p \times p$ matrix.*

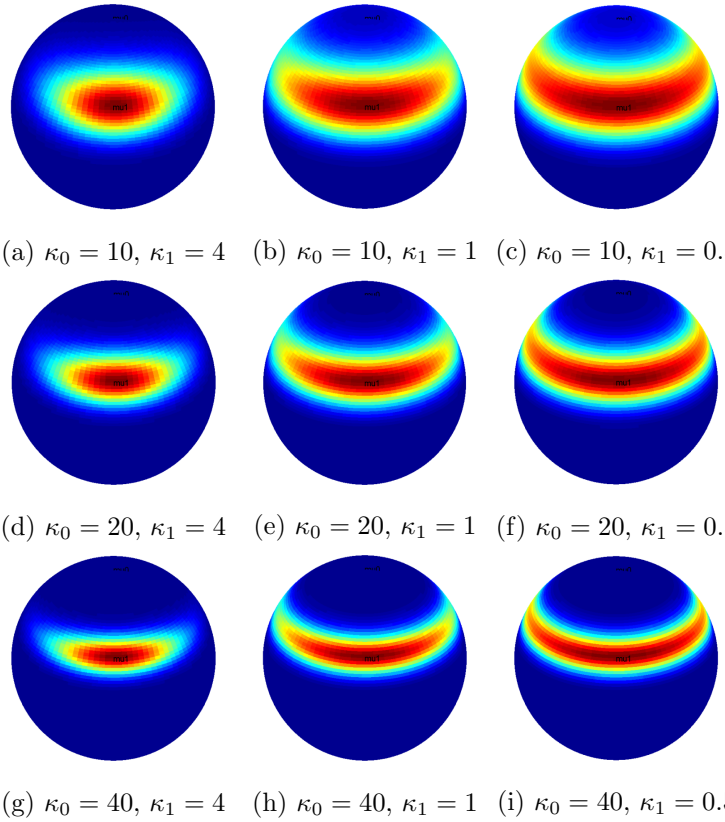
*(i) If $B\mu_0 = \mu_0$ and $B\mu_1 = \mu_1$ then $X$ and $BX$ have the same distribution and so do $Y$ and $BY$.*

*(ii) If $B$ is orthogonal and $X, BX$ have the same distribution (or $Y, BY$ have the same distribution), then $B\mu_0 = \mu_0$ and $B\mu_1 = \mu_1$.*

*(iii) $X \sim S1(-\mu_0, \mu_1, \kappa_0, \kappa_1)$ and $Y \sim S2(-\mu_0, \mu_1, \kappa_0, \kappa_1)$.*

An example for the matrix $B$ in Proposition 1(i) is the reflection matrix $B = I_p - 2UU^\top$, where $U = [u_3, \ldots, u_p]$ is such that $[u_1, \ldots, u_p]$ is a $p \times p$ orthogonal matrix whose column vectors $u_1$ and $u_2$ generate $\mu_0$ and $\mu_1$.

*Remark* 1. The S1 distribution is a special case of the Fisher-Bingham distribution (Mardia, 1975). Following the notation of Kent (1982), the S1 distribution may be labeled as a $FB_6$ *distribution*, in the special case of $p = 3$, emphasizing the 6-dimensional parameter space. In



(a) $\kappa_0 = 10$, $\kappa_1 = 4$ (b) $\kappa_0 = 10$, $\kappa_1 = 1$ (c) $\kappa_0 = 10$, $\kappa_1 = 0.5$

(d) $\kappa_0 = 20$, $\kappa_1 = 4$ (e) $\kappa_0 = 20$, $\kappa_1 = 1$ (f) $\kappa_0 = 20$, $\kappa_1 = 0.5$

(g) $\kappa_0 = 40$, $\kappa_1 = 4$ (h) $\kappa_0 = 40$, $\kappa_1 = 1$ (i) $\kappa_0 = 40$, $\kappa_1 = 0.5$

Figure 2: The S1 densities on $\mathbb{S}^2$ modeling non-isotropic small-circle distributions. High density (red), low density (blue). In all figures, $\mu_0$ points to the north pole and $\mu_1$ satisfies $\mu_0^\top \mu_1 = 1/2$. Rows and columns correspond to different choices of concentration parameters ($\kappa_0$, $\kappa_1$).

terms of the general parameterization of the Fisher-Bingham density (cf. Mardia and Jupp, 2000, p.174), we write $\gamma = 2\kappa_0\nu\mu_0 + \kappa_1\mu_1$ and $A = \kappa_0\mu_0\mu_0^\top$, so that the S1 density (5) is expressed as

$$f(x;\gamma,A) = \frac{1}{\alpha(\gamma,A)} \exp\{\gamma^\top x - x^\top A x\}, \tag{10}$$

where

$$\alpha(\gamma,A) = a(\kappa_0,\kappa_1,\nu) \exp\left\{\kappa_0\nu^2\right\}. \tag{11}$$

This relation to the general Fisher-Bingham distribution facilitates a random data generation and maximum likelihood estimation, shown later in Sections 2.4 and 3.1.

### 2.3 Multivariate extensions

The univariate small-sphere distributions (5–6) are now extended to model a tuple of *associated* random directions, $\mathbf{X} = (X_{(1)},\ldots,X_{(K)}) \in (\mathbb{S}^{p-1})^K$. We confine ourselves to a special case where the marginal distributions of $X_{(k)}$ are all either S1 or S2, with a common "axis" parameter

$\mu_0$. We begin by introducing multivariate small-sphere distributions for independent random directions.

**Independent extensions.** Suppose that, in the $K$-tuple of random directions $\mathbf{X}$, each $X_{(k)} \in \mathbb{S}^{p-1}$ is marginally distributed as S1$(\mu_0, \mu_{1(k)}, \kappa_{0(k)}, \kappa_{1(k)})$. Throughout, we assume that $\nu_{(k)} = \mu_0^\top \mu_{1(k)} \in (-1, 1)$ so that the underlying small spheres do not degenerate. If components of $\mathbf{X}$ are mutually independent, then the joint density evaluated at $\mathbf{x} \in (\mathbb{S}^{p-1})^K$ is proportional to

$$\exp\left\{\Gamma^\top \mathbf{x} - \mathbf{x}^\top \mathbf{A} \mathbf{x}\right\}. \tag{12}$$

Here, $\Gamma = [\gamma_{(1)}, \ldots, \gamma_{(K)}]^\top$, where $\gamma_{(k)} = 2\kappa_{0(k)}\nu_{(k)}\mu_0 + \kappa_{1(k)}\mu_{1(k)}$, and $\mathbf{A} = D_{\kappa_0} \otimes (\mu_0 \mu_0^\top)$, where $D_{\kappa_0} = \mathrm{diag}(\kappa_{0(1)}, \ldots, \kappa_{0(K)})$. Each marginal density is of the form (10).

If each component is marginally distributed as S2$(\mu_0, \mu_{1(k)}, \kappa_{0(k)}, \kappa_{1(k)})$, then writing the density in terms of $(s, y)$ as done for (9) facilitates our discussion. First, we decompose each $x_{(k)}$ into $s_{(k)} = \mu_0^\top x_{(k)} \in [-1, 1]$ and $y_{(k)} \in \mathbb{S}^{p-2}$ as defined in (7). Further, we denote $\widetilde{\mu_{1(k)}}$ for the scaled projection of $\mu_{1(k)}$ as done for the univariate case. Then an independent multivariate extension for the S2 model can be expressed as the joint density of $\mathbf{s} = (s_{(1)}, \ldots, s_{(K)})$ and $\mathbf{y} = (y_{(1)}, \ldots, y_{(K)})$, proportional to

$$\exp\left\{H^\top \mathbf{s} - \mathbf{s}^\top D_{\kappa_0} \mathbf{s} + \mathbf{M}^\top \mathrm{vec}(\mathbf{y})\right\}, \tag{13}$$

where $H = (2\kappa_{0(1)}\nu_{(1)}, \ldots, 2\kappa_{0(K)}\nu_{(K)})$ and $\mathbf{M} = \mathrm{vec}(\kappa_{1(1)}\widetilde{\mu_{1(1)}}, \ldots, \kappa_{1(K)}\widetilde{\mu_{1(K)}})$ while $\mathrm{vec}(\cdot)$ denotes the usual vectorization of a matrix.

In the following, we denote the independent multivariate model of S1 by *iMS1* and the independent multivariate model of S2 by *iMS2*.

**Vertical and horizontal dependence.** Based on (12) and (13), we now contemplate on dependent models. Obviously, if we allow in (12) nonzero offdiagonal entries of $\mathbf{A}$, then we obtain a dependent modification of the S1 model. With our applications in mind, however, we aim at modeling a specific structure of dependence that is natural to the variables $(\mathbf{s}, \mathbf{y})$ in (13).

If $s_{(1)}, \ldots, s_{(K)}$ are dependent, we speak of *vertical dependence*; if $y_{(1)}, \ldots, y_{(K)}$ are dependent, we speak of *horizontal dependence*. In practice, when we deal with small-circle concentrated directional data, association among these vectors usually occurs along small-circles with independent vertical errors. For example, when a 3D object is modeled by skeletal representations, as described in more detail in Section 6 and visualized in Fighre 6, a deformation of the object is measured by the movements of directional vectors on $\mathbb{S}^2$. When a simple rotational deformation

(such as bending, twisting or rotation) occurs, all the directions move along small-circles with a common axis $\mu_0$. In this situation, the longitudinal variations along the circles are dependent to each other because nearby spoke vectors are under the effect of similar deformations. (Examples of such longitudinal dependence can be found in Section 6 as well as in Schulz et al. (2015).) Adding such a horizontal (or longitudinal) dependence to a multivariate S1 model requires a careful parametrization of the offdiagonal entries of $\mathbf{A}$ in (12). This has not been straightforward, and we leave it for future work. On the other hand, it is feasible to extend the S2 model by generalizing the "vMF part" of $\mathbf{y}$, the last term in the exponent of (13), to a Fisher-Bingham type.

To this end, we introduce a parameter matrix $\mathbf{B}$ to model general quadratics in $\mathrm{vec}(\mathbf{y})$. This allows to write the densities for a general multivariate small-sphere distribution of the second kind (GMS2) as follows:

$$\begin{aligned} & g_{\mathrm{GMS2}}(\mathbf{s}, \mathbf{y}; H, D_{\kappa_0}, \mathbf{M}, \mathbf{B}) \\ &= \frac{1}{T(H, D_{\kappa_0}, \mathbf{M}, \mathbf{B})} \exp\left\{H^\top \mathbf{s} - \mathbf{s}^\top D_{\kappa_0} \mathbf{s} + \mathbf{M}^\top \mathrm{vec}(\mathbf{y}) + \mathrm{vec}(\mathbf{y})^\top \mathbf{B}\, \mathrm{vec}(\mathbf{y})\right\} \end{aligned} \tag{14}$$

where $H, D_{\kappa_0}$ and $\mathbf{M}$ as defined in (13), and $T(H, \kappa_0, \mathbf{M}, \mathbf{B})$ is a normalizing constant. To ensure the marginal distributions of $(s_{(k)}, y_{(k)})$ is a S2 distribution of the form (9), we set

$$\mathbf{B} = (B_{k,l})_{k,l=1}^{K},\ B_{k,l} = (b_{i,j}^{(k,l)})_{i,j=1}^{p-1}$$

as a block matrix with vanishing blocks $B_{k,k} = 0$ on the diagonal. The submatrix $B_{k,l}$ models the horizontal association between $y_{(k)}$ and $y_{(l)}$. The fact that $z^\top \mathbf{B} z = z^\top \mathbf{B}^\top z = \frac{1}{2} z^\top (\mathbf{B} + \mathbf{B}^\top) z$ for any vector $z \in \mathbb{R}^{(p-1)K}$ allows us to assume without loss of generality that $\mathbf{B}$ is symmetric.

**An MS2 distribution on $(\mathbb{S}^2)^K$.** As a viable submodel for the practically important case $p = 3$, we propose to use a special form for the offdiagonal blocks $B_{k,l}$ of $\mathbf{B}$. In particular, with $\lambda_{k,l}$ representing the degrees of association between $y_{(k)}$ and $y_{(l)}$, we set

$$\begin{aligned} B^{k,l} &= 2 \begin{pmatrix} \widetilde{\mu_{1(k)}} & \widetilde{\mu_{2(k)}} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \lambda_{k,l} \end{pmatrix} \begin{pmatrix} \widetilde{\mu_{1(l)}} & \widetilde{\mu_{2(l)}} \end{pmatrix}^\top \\ &= 2\lambda_{k,l}\, \widetilde{\mu_{2(k)}} \widetilde{\mu_{2(l)}}^\top, \end{aligned} \tag{15}$$

where $\begin{pmatrix} \widetilde{\mu_{1(k)}} & \widetilde{\mu_{2(k)}} \end{pmatrix}$ is the rotation matrix given by setting

$$\widetilde{\mu_{2(k)}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \widetilde{\mu_{1(k)}}.$$

The density (14) with the above parameterization of $\mathbf{B}$ will be referred to as a multivariate S2 distribution (MS2) for data on $(\mathbb{S}^2)^K$.

Our choice of the simple parametrization (15) does not restrict the modeling capability of the general model (14), and has some advantages in parameter interpretations and also in estimation. To see this, we resort to use an angle representation for $\mathbf{y}$ (available to this $p = 3$ case). For each $k$, define $\phi_{(k)}$ and $\zeta_{(k)}$ such that $y_{(k)} = (\cos\phi_{(k)}, \sin\phi_{(k)})^\top$ and $\widetilde{\mu_{1(k)}} = (\cos\zeta_{(k)}, \sin\zeta_{(k)})^\top$. Accordingly, the inner products appearing in (14) can be expressed by

$$\widetilde{\mu_{1(k)}}^\top y_{(k)} = \cos(\phi_{(k)} - \zeta_{(k)}), \quad \widetilde{\mu_{2(k)}}^\top y_{(k)} = \sin(\phi_{(k)} - \zeta_{(k)}).$$

Let $\boldsymbol{\phi} = (\phi_{(1)}, \ldots, \phi_{(K)})^\top$, $\boldsymbol{\zeta} = (\zeta_{(1)}, \ldots, \zeta_{(K)})^\top$, $\boldsymbol{\kappa}_1 = (\kappa_{1(1)}, \ldots, \kappa_{1(K)})^\top$,

$$\begin{aligned} c(\boldsymbol{\phi}, \boldsymbol{\zeta}) &= \big(\cos(\phi_{(1)} - \zeta_{(1)}), \ldots, \cos(\phi_{(K)} - \zeta_{(K)})\big)^\top, \\ s(\boldsymbol{\phi}, \boldsymbol{\zeta}) &= \big(\sin(\phi_{(1)} - \zeta_{(1)}), \ldots, \sin(\phi_{(K)} - \zeta_{(K)})\big)^\top, \end{aligned}$$

and $\boldsymbol{\Lambda} = (\lambda_{k,l})_{k,l=1}^K$ where $\lambda_{k,l} (= \lambda_{l,k})$ for $k \neq l$ is the association parameter used in (15), and $\lambda_{k,k}$ is set to zero. The density of the MS2 distribution, in terms of $(\mathbf{s}, \boldsymbol{\phi})$, is then

$$\begin{aligned} &g_{\mathrm{MS2}}(\mathbf{s}, \boldsymbol{\phi}; H, D_{\kappa_0}, \boldsymbol{\kappa}_1, \boldsymbol{\zeta}, \boldsymbol{\Lambda}) \\ &= \frac{1}{T(H, D_{\kappa_0}, \boldsymbol{\kappa}_1, \boldsymbol{\zeta}, \boldsymbol{\Lambda})} \exp\left\{ H^\top \mathbf{s} - \mathbf{s}^\top D_{\kappa_0} \mathbf{s} + \boldsymbol{\kappa}_1^\top c(\boldsymbol{\phi}, \boldsymbol{\zeta}) + \frac{1}{2} s(\boldsymbol{\phi}, \boldsymbol{\zeta})^\top \boldsymbol{\Lambda} s(\boldsymbol{\phi}, \boldsymbol{\zeta}) \right\}. \quad (16) \end{aligned}$$

From (16), it can be easily seen that the "horizontal angles" $\boldsymbol{\phi}$ follow the *multivariate von Mises* distribution (Mardia et al., 2008) and are independent to the vertical component $\mathbf{s}$. As we will see later in Section 3.2, this facilitates estimations for the MS2 distributions. Moreover, since

$$\begin{aligned} &\boldsymbol{\kappa}_1^\top c(\boldsymbol{\phi}, \boldsymbol{\zeta}) + \frac{1}{2} s(\boldsymbol{\phi}, \boldsymbol{\zeta})^\top \boldsymbol{\Lambda} s(\boldsymbol{\phi}, \boldsymbol{\zeta}) \\ &= \sum_{k=1}^K \kappa_{1(k)} \left(1 - \frac{(\phi_k - \zeta_k)^2}{2}\right) + \frac{1}{2} \sum_{k=1}^K \sum_{k \neq l = 1}^K \Big(\lambda_{kl} (\phi_k - \zeta_k)(\phi_l - \zeta_l)\Big) + o(\|\boldsymbol{\phi} - \boldsymbol{\zeta}\|^2), \quad (17) \end{aligned}$$

for large enough concentrations, $\boldsymbol{\phi}$ is approximately multivariate normal with mean $\boldsymbol{\zeta}$ and a precision matrix $\Sigma^{-1}$, where $(\Sigma^{-1})_{kk} = \kappa_{1(k)}$ and $(\Sigma^{-1})_{kl} = -\lambda_{kl}$ for $1 \leq k \neq l \leq K$. These parameters are naturally interpreted as partial variances and correlations. This interpretation of the parameters as entries of a precision matrix is most immediate under the MS2, but is not under the general case.

### 2.4 Random data generation

Generating pseudo-random samples from the S1 and S2 distributions are important in simulations and in developments of computer-intensive inference procedures.

For random sampling from the S1 (5) or the iMS1 (12) distribution , the fact that each marginal distribution of the iMS1 is a special case of the Fisher-Bingham is handy. Thereby, one can use the Markov Chain Monte Carlo algorithm of Hoff (2009), who provided detailed procedures of generating Fisher-Bingham-variate samples (10).

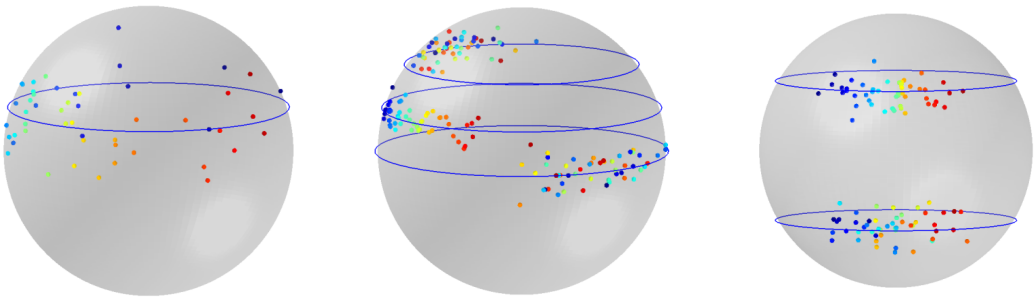
For random sampling from the S2 (6) or MS2 (16) distribution, we take advantage of the independence between the pair $(\mathbf{s}, \mathbf{y})$. As we assume vertical independence (i.e., $s_{(1)}, \ldots, s_{(K)}$ are independent), each $s_{(k)}$ can be sampled separately. Therefore, sampling from the MS2 distribution amounts to independently drawing samples from a truncated normal distribution (for $s_{(k)}$) and from a multivariate von Mises distribution (for $\mathbf{y}$). Specifically, to sample $\mathbf{x} = (x_{(1)}, \ldots, x_{K)})$ from MS2 with parameters $(\mu_0, \boldsymbol{\mu}_1, \boldsymbol{\kappa}_0, \boldsymbol{\kappa}_1, \boldsymbol{\Lambda})$, the following procedure can be used.

Step 1. For each $k$, sample $s_{(k)}$ from the truncated normal distribution with mean $\nu_{(k)}$ and variance $1/(2\kappa_{0(k)})$, truncated to the interval (-1,1).

Step 2. For the S2 or iMS2 model, sample each $y_{(k)} \in \mathbb{S}^{p-2}$ in $\mathbf{y} = (y_{(1)}, \ldots, y_{(K)})$ independently from the von Mises distribution with mean $(1, 0, \ldots, 0)$ and concentration $\kappa_{1(k)}$; for the MS2 distribution (when $p = 3$), sample the $K$-tuple $\mathbf{y} \in (\mathbb{S}^1)^K$ directly from the multivariate von-Mises distribution with mean $(1, 0)$ and precision parameters $\boldsymbol{\kappa}_1$ and $\boldsymbol{\Lambda}$.

Step 3. For each $k$, let $E_{(k)}$ be a $p \times p$ orthogonal matrix with $(\mu_0, P_{\mu_0}\mu_{1(k)})$ being the first two column vectors. Set $x_{(k)} = E_{(k)}^{\top}\left(s_{(k)}, (1 - s_{(k)}^2)^{-1/2} y_{(k)}\right)$.

In our experiments, sampling from the S2 and MS2 distributions is much faster than from the S1. In particular, when the dimension $p$ or the concentration level is high, the Markov Chain simulations for the S1 appear to be sluggish. Some examples of random samples from the S2, iMS2 and MS2 distributions are shown in Fig. 3. The small-circles $C(\mu_0, \nu_{(k)})$ are also overlaid in the figure. In particular, the MS2 sample in the rightmost panel clearly show a horizontal dependence.



(a) S2 on $\mathbb{S}^2$ (b) iMS2 on $(\mathbb{S}^2)^3$ (c) MS2 on $(\mathbb{S}^2)^2$

Figure 3: Random samples from the S2, iMS2 and MS2 distributions. Same colors represent same observations for (a) low concentrations ($\kappa_0 = 10$, $\kappa_1 = 1$), (b) independent directions with high concentrations ($\kappa_{0(k)} = 100$, $\kappa_{1(k)} = 10$, $k = 1, 2, 3$), (c) horizontally dependent directions with high concentrations ($\kappa_{0(k)} = 50$, $\kappa_{1(k)} = 30$) and high dependence ($\lambda_{12} = 24$).

# 3 Estimation

## 3.1 Approximate maximum likelihood estimation for the S1

The standard way to estimate parameters of the S1 is to use the maximum likelihood estimates (m.l.e). However, it does not seem possible to obtain explicit expressions of the m.l.e., partly due to having no closed-form expression of the normalizing constant (5). We propose to use an approximate m.l.e., obtained by iterating between updates for $\mu_0$ and for $(\mu_1, \kappa_0, \kappa_1)$, where each separate problem can be solved efficiently. The proposed estimation procedure, given below, only needs to specify an initial value for $\nu = \mu_0^\top \mu_1 \in (-1, 1)$. Our procedure naturally extends to estimation for the independent multivariate S1 distribution, which will also be discussed.

As a preparation, we first describe an approximation of the normalizing constant. For this, we follow Kume and Wood (2005), who used saddlepoint density approximations for approximating normalizing constants of Fisher-Bingham distributions. Since the proposed S1 distribution has not been discussed in Kume and Wood (2005) nor in their later work (Kume et al., 2013), we describe an application of the approximations to the S1 model.

The normalizing constant $a(\kappa_0, \kappa_1, \nu)$ has an alternative expression, as shown in the following.

**Proposition 2.** *For any $h > 0$, let $\xi = (\frac{\nu(2\kappa_0+\kappa_1)}{2(\kappa_0+h)}, \frac{\kappa_1\sqrt{1-\nu^2}}{2h}, 0, \ldots, 0)^\top \in \mathbb{R}^p$ and $\Psi$ be the $p \times p$ diagonal matrix with diagonal elements $(\kappa_0 + h, h, \ldots, h)$. Moreover, let $g(r)$ $(r > 0)$ be the probability density function of $R = Z^\top Z$, where $Z \sim N_p(\xi, \frac{1}{2}\Psi^{-1})$. Then the normalizing constant $a(\kappa_0, \kappa_1, \nu)$ of (5) is*

$$a(\kappa_0, \kappa_1, \nu) = 2\pi^{p/2}|\mathbf{\Psi}|^{-1/2} g(1) \exp\left(\xi^T \Psi \xi + h - \kappa_0 \nu^2\right). \tag{18}$$

In Proposition 2, the function $g$ is the density of a linear combination of independent non-central $\chi^2_1$ random variables. Kume and Wood (2005) showed a use of saddle-point density approximations in the numerical computation of $g(1)$. Following Kume and Wood, we first note that the derivatives of the cumulant generating function, $K_g(t) = \log \int_0^\infty e^{tr} g(r)dr$, associated with the density $g$ have closed-form expressions. Denote by $K_g^{(j)}(t)$ the $j$th derivative of $K_g(t)$, for $j = 1, \ldots, 4$, we get

$$K_g^{(j)}(t) = \frac{(j-1)!}{2}\left(\frac{1}{(\kappa_0 + h - t)^j} + \frac{p-1}{(h-t)^j}\right) + \frac{j!}{4}\left(\frac{\nu^2(2\kappa_0+\kappa_1)^2}{(\kappa_0+h-t)^{j+1}} + \frac{\kappa_1^2(1-\nu^2)}{(h-t)^{j+1}}\right).$$

Let $\hat{t}$ be the unique solution in $(-\infty, h)$ of the saddle-point equation $K_g^{(1)}(t) = 1$, which can be easily evaluated by using, e.g., a bisection method. Then a saddle-point density approximation of $g(1)$ is

$$\hat{g}(1) = (2\pi K_g^{(2)}(\hat{t}))^{-1/2} \exp(K_g(\hat{t}) - \hat{t} + T), \tag{19}$$

where $T = K_g^{(4)}(\hat{t})/\{8(K_g^{(2)}(\hat{t}))^2\} - 5(K_g^{(3)}(\hat{t}))^2/\{24(K_g^{(2)}(\hat{t}))^3\}$. In the following, we approximate the value of $a(\kappa_0, \kappa_1, \nu)$ by $\hat{a}(\kappa_0, \kappa_1, \nu)$ obtained by plugging (19) in place of $g(1)$ in (18).

We are now ready to describe our estimating procedure. Suppose $x_1, \ldots, x_n$ is a sample from $\mathrm{S1}(\mu_0, \mu_1, \kappa_0, \kappa_1)$. Given an initial value $\hat{\nu}^{(0)} \in (-1, 1)$, we iterate between steps 1 and 2 below until convergence.

**Step 1**: Updating $\mu_0$ given all other parameters, of which we only need $\nu$. Suppose the inner product $\nu = \mu_0^\top \mu_1 \in (-1, 1)$ is fixed. Then maximizing the likelihood function with respect to $\mu_0$ is equivalent to minimizing $\frac{1}{n}\sum_{i=1}^n (\mu_0^\top x_i - \nu)^2$ subject to the constraint $\mu_0^\top \mu_0 = 1$. With a Lagrangian multiplier $\lambda$ and using a matrix notation, we solve

$$\min_{\mu_0} \frac{1}{n}\|\mathbb{X}^\top \mu_0 - \nu 1_n\|^2 - \lambda(\mu_0^\top \mu_0 - 1), \tag{20}$$

where $\mathbb{X}$ is the $p \times n$ matrix whose $i$th column is $x_i$. This problem (20) is a special case of quadratically constrained quadratic minimization problems (Albers et al., 2011). The first order condition of (20) leads to

$$S\mu_0 - \nu\bar{x} - \lambda\mu_0 = 0,$$

where $S = \mathbb{X}\mathbb{X}^\top/n$, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$. For a fixed Lagrangian multiplier $\lambda$, the solution is $\hat{\mu}_0 = \nu(S - \lambda I_p)^{-1}\bar{x}$, provided that $S$ is of full rank. The constraint $\mu_0^\top \mu_0 = 1$ makes us find a root $\lambda$ of $\nu^2 \bar{x}^\top (S - \lambda I_p)^{-2}\bar{x} - 1$. The root $\hat{\lambda}$ is found by a bisection search in the range $[-\nu^2\bar{x}^\top\bar{x}, \lambda_S]$, where $\lambda_S > 0$ is the smallest eigenvalue of $S$ (Browne, 1967; Albers et al., 2011). The solution to (20) is then

$$\hat{\mu}_0 = \nu(S - \hat{\lambda} I_p)^{-1}\bar{x}. \tag{21}$$

If $\nu = 0$, then $\hat{\mu}_0$ is the eigenvector of $S$ corresponding to the smallest eigenvalue.

**Step 2**: Updating $(\mu_1, \kappa_0, \kappa_1)$, given $\mu_0$. To facilitate the estimation of $\mu_1$, we define the $E(X) \in \mathbb{R}^p$ as the usual mean of random vector $X$ in the ambient space $\mathbb{R}^p$. We use the fact that $\mu_1$ is a linear combination of $\mu_0$ and $\gamma_0 := E(X)/\|E(X)\|$ (see Lemma 3 in Appendix) and reparameterize $\mu_1$ by an angle $\varphi$ and a direction $\gamma^*$ orthogonal to $\mu_0$. By writing $\gamma^* = P_{\mu_0}\gamma_0/\|P_{\mu_0}\gamma_0\|$,

$$\mu_1 = \cos(\varphi)\mu_0 + \sin(\varphi)\gamma^*. \tag{22}$$

With $\gamma_0$ estimated by $\hat{\gamma}_0 = \bar{x}/\|\bar{x}\|$, we now optimize for $\varphi \in [0, 2\pi)$ together with $\kappa_0, \kappa_1$. With $\mu_0$ and $\hat{\gamma}_0$ (thus $\hat{\gamma}^*$) given, the approximate negative log-likelihood with respect to $(\varphi, \kappa_0, \kappa_1)$ is

$$\tilde{\ell}_{\mu_0}(\varphi, \kappa_0, \kappa_1) = -n \log \hat{a}(\kappa_0, \kappa_1, \varphi) - \kappa_0 \sum_{i=1}^{n} (\mu_0^\top x_i - \cos\varphi)^2 + \kappa_1 \sum_{i=1}^{n} x_i^\top (\cos(\varphi)\mu_0 + \sin(\varphi)\hat{\gamma}^*). \tag{23}$$

Numerically minimizing (23) is much simpler than optimizing for $\mu_1$ with the nonlinear constraint $\|\mu_1\| = 1$. We use a standard optimization package to obtain $\hat{\varphi}, \hat{\kappa}_0, \hat{\kappa}_1$ that minimizes (23). We get $\hat{\mu}_1$ by substituting $(\gamma^*, \varphi)$ by $(\hat{\gamma}^*, \hat{\varphi})$ in (22) and $\hat{\nu} = \cos(\hat{\varphi})$.

Let us now describe an extension of the above algorithm to the iMS1 model. Suppose $(x_{i(1)}, \ldots, x_{i(K)}) \in (S^{p-1})^K$ for $i = 1, \ldots, n$ is a sample from an iMS1 model, where each marginal distribution is $\mathrm{S1}(\mu_0, \mu_{1(j)}, \kappa_{0(j)}, \kappa_{1(j)})$. While Step 2 above can be applied to update $\mu_{1(k)}, \kappa_{0(k)}, \kappa_{1(k)}$ given $\mu_0$, we modify Step 1 by replacing (20) with

$$\min_{\mu_0} \frac{1}{n} \sum_{j=1}^{K} (\kappa_{0(j)} \|\mathbb{X}_{(j)}^\top \mu_0 - \nu_{(j)} 1_n\|^2) - \lambda(\mu_0^\top \mu_0 - 1),$$

where the marginal $p \times n$ observation matrix $\mathbb{X}_{(j)}$ consists of $x_{i(j)}$ $(i = 1, \ldots, n)$, which can be solved in a similar fashion to (21).

In our experiments, choosing a good initial value of $\hat{\nu}^{(0)}$ is imperative in ensuring fast convergence to the global maximum. We discuss several choices of initial values in the online supplementary material.

### 3.2 Estimation via profile likelihood for the S2

The S2 model has the convenient independence property among the horizontal and vertical components. To take advantage of this, suppose for now that $\mu_0$ is known. This allows to decompose an observation $x$ into two independent random variables $(s, y)$, which in turn leads to an easy estimation of the remaining parameters $\eta := (\mu_1, \kappa_0, \kappa_1)$. Thus our strategy of computing the m.l.e. is in two nested steps. Let $\ell_n(\mu_0, \eta)$ be the negative likelihood function

given a sample $x_1, \ldots, x_n$ from S2$(\mu_1, \eta)$. In the outer step, we update $\mu_0$ to maximize a profile likelihood, i.e.,

$$\hat{\mu}_0 = \underset{\mu_0}{\operatorname{argmin}}\, \ell_n(\mu_0, \hat{\eta}_{\mu_0}), \tag{24}$$

where evaluating

$$\hat{\eta}_{\mu_0} = \underset{\eta}{\operatorname{argmin}}\, \ell_n(\mu_0, \eta) \tag{25}$$

for a fixed $\mu_0$ is the inner step. It is straightforward to see that the m.l.e. of $(\mu_0, \eta)$ is given by $(\hat{\mu}_0, \hat{\eta}_{\hat{\mu}_0})$.

In the following, we discuss in detail the inner step (25) of minimizing $\ell_{\mu_0}(\eta) := \ell_n(\mu_0, \eta)$ for the independent multivariate S2 model (13) and for the MS2 model (16), while we resort to use a standard optimization package for solving (24). For simplicity, we focus on the $p = 3$ case, but an extension to $p > 3$ is straightforward in the independent case.

**Independent multivariate S2.** Suppose $(x_{i(1)}, \ldots, x_{i(K)}) \in (S^{p-1})^K$ for $i = 1, \ldots, n$ is a sample from an independent multivariate S2, where each marginal distribution is S2$(\mu_0, \mu_{1(j)}, \kappa_{0(j)}, \kappa_{1(j)})$. For a given $\mu_0$, the joint density can be written in terms of $(\mathbf{s}_i, \boldsymbol{\phi}_i)$ as done in (16), but with $\boldsymbol{\Lambda} = 0$. Furthermore, with $\boldsymbol{\kappa}_0 = (\kappa_{0(1)}, \ldots, \kappa_{0(K)})^\top$, $\boldsymbol{\nu} = (\nu_{(1)}, \ldots, \nu_{(K)})^\top$, we can write

$$H^\top \mathbf{s}_i - \mathbf{s}_i^\top D_{\kappa_0} \mathbf{s}_i = -(\boldsymbol{\kappa}_0 \circ (\mathbf{s}_i - \boldsymbol{\nu}))^\top (\mathbf{s}_i - \boldsymbol{\nu}),$$

where $\circ$ denotes the element-wise product, and

$$\log T(H, D_{\kappa_0}, \boldsymbol{\kappa}_1, \boldsymbol{\zeta}, 0) = \sum_{j=1}^{K} \left[ \log b(\kappa_{0(j)}, \kappa_{1(j)}, \nu_{(j)}) + \kappa_{0(j)} \nu_{(j)}^2 \right]. \tag{26}$$

Note that the normalizing constant $b(\kappa_0, \kappa_1, \nu)$ satisfies

$$\begin{aligned} b(\kappa_0, \kappa_1, \nu) &= \int_0^{2\pi} \exp(\kappa_1 \cos(\phi - \zeta)) d\phi \int_{-1}^{1} \exp\{-\kappa_0 (s - \nu)^2\} ds \\ &= (2\pi)^{3/2} (2\kappa_0)^{-1/2} \mathcal{I}_0(\kappa_1) \left[ \Phi((1-\nu)\sqrt{2\kappa_0}) - \Phi(-(1+\nu)\sqrt{2\kappa_0}) \right], \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Finally, the negative log-likelihood function (given $\mu_0$) is

$$\ell_{\mu_0}(\boldsymbol{\nu}, \boldsymbol{\zeta}, \boldsymbol{\kappa}_0, \boldsymbol{\kappa}_1; \{\mathbf{s}_i, \boldsymbol{\phi}_i\}_{i=1}^n) = \ell_{\mu_0}^{(1)}(\boldsymbol{\nu}, \boldsymbol{\kappa}_0) + \ell_{\mu_0}^{(2)}(\boldsymbol{\zeta}, \boldsymbol{\kappa}_1), \tag{27}$$

where

$$\ell^{(1)}_{\mu_0}(\boldsymbol{\nu},\boldsymbol{\kappa}_0) = \sum_{j=1}^{K}\left[\kappa_{0(j)}\sum_{i=1}^{n}(s_{i(j)}-\nu_{(j)})^2 - \frac{n}{2}\log(2\kappa_{0(j)}) + \frac{n}{2}\log(2\pi)\right. \tag{28}$$
$$\left. + n\log\left(\Phi\left((1-\nu_{(j)})\sqrt{2\kappa_{0(j)}}\right) - \Phi\left(-(1+\nu_{(j)})\sqrt{2\kappa_{0(j)}}\right)\right)\right],$$
$$\ell^{(2)}_{\mu_0}(\boldsymbol{\zeta},\boldsymbol{\kappa}_1) = -\sum_{j=1}^{K}\left[\kappa_{1(j)}\sum_{i=1}^{n}\cos(\phi_{i(j)}-\zeta_{(j)}) - n\log\mathcal{I}_0(\kappa_{1(j)}) - n\log(2\pi)\right]. \tag{29}$$

Therefore, the optimization for the inner step (25) is equivalent to simultaneously solving $2K$ subproblems.

Each of the $K$ subproblems of (28) is equivalent to obtaining the m.l.e. of a truncated normal distribution $\mathrm{trN}(\nu_{(j)},(2\kappa_{0(j)})^{-1/2};(-1,1))$ based on the observations $s_{i(j)}$ $(i=1,\ldots,n)$. Similarly, each subproblem of (29) amounts to obtaining the m.l.e. of a von Mises distribution with mean $\zeta_{(j)}$ and concentration $\kappa_{1(j)}$ from the sample $\phi_{i(j)}$ $(i=1,\ldots,n)$. The m.l.e.s of the truncated normal are numerically computed, and we use the method of Banerjee et al. (2005) to obtain approximations of the m.l.e.s of the von Mises.

**MS2.** Under the general MS2 model (16) with a dependence structure on $\boldsymbol{\phi}_i$, the decomposition $\ell_{\mu_0}(\boldsymbol{\nu},\boldsymbol{\zeta},\boldsymbol{\kappa}_0,\boldsymbol{\kappa}_1,\boldsymbol{\Lambda}) = \ell^{(1)}_{\mu_0}(\boldsymbol{\nu},\boldsymbol{\kappa}_0) + \ell^{(2)}_{\mu_0}(\boldsymbol{\zeta},\boldsymbol{\kappa}_1,\boldsymbol{\Lambda})$, similar to (27), is still valid, except that (29) is replaced by

$$\ell^{(2)}_{\mu_0}(\boldsymbol{\zeta},\boldsymbol{\kappa}_1,\boldsymbol{\Lambda}) = -\sum_{i=1}^{n}\left[\boldsymbol{\kappa}_1^\top c(\boldsymbol{\phi}_i,\boldsymbol{\zeta}) + \frac{1}{2}s(\boldsymbol{\phi}_i,\boldsymbol{\zeta})^\top\boldsymbol{\Lambda}s(\boldsymbol{\phi}_i,\boldsymbol{\zeta}) - \log T(\boldsymbol{\kappa}_1,\boldsymbol{\nu},\boldsymbol{\Lambda})\right]. \tag{30}$$

Minimizing (30) is equivalent to computing the m.l.e. of the multivariate von Mises (Mardia et al., 2008). We either use maximum pseudo-likelihood estimate as discussed in Mardia et al. or the method of moment estimates, defined as

$$\hat{\zeta}_{(j)} = \frac{1}{n}\sum_{i=1}^{n}\phi_{i(j)}/\|\frac{1}{n}\sum_{i=1}^{n}\phi_{i(j)}\|, \quad \hat{\kappa}_{1(j)} = \bar{S}^{-1}_{jj}, \quad \hat{\lambda}_{(jk)} = \bar{S}^{-1}_{jk} \quad (j\neq k), \tag{31}$$

where $\bar{S} = (\bar{S}_{jk})$ and $\bar{S}_{jk} = \frac{1}{n}\sum_{i=1}^{n}\sin(\phi_{i(j)}-\hat{\zeta}_{(j)})\sin(\phi_{i(k)}-\hat{\zeta}_{(k)})$ for $j,k=1,\ldots,K$. These method of moment estimates coincide with the m.l.e.s when $K=2$. For larger $K>3$, the accuracy of the method of moment estimates deteriorates, but evaluating m.l.e.s or a maximum pseudo-likelihood estimates becomes computationally prohibitive.

# 4 Testing hypotheses

It is of interest to make inference on some of the parameters. In this section, we describe a large-sample testing procedure for several hypotheses of interest.

Our testing procedure is based on the likelihood ratio statistic, with effort devoted to an identification of the restricted parameter space, $\Theta_0$, for each hypothesis and computing maximized likelihood under $\Theta_0$. For simplicity, we describe our procedure using the S1 distribution. The test procedures using the iMS1, iMS2, or MS2 distributions can be similarly developed, and we omit the details here. Denote the parameter space of the S1 distribution by $\Theta = (\mathbb{S}^{p-1})^2 \times (\mathbb{R}_+)^2$. For some $\Theta_0$ that dictates a null hypothesis $H_0$ and satisfies $\Theta_0 \subset \Theta$, we denote the maximized log-likelihood under $\Theta_0$ by $\mathcal{L}_0$, and the maximized log-likelihood under $\Theta$ by $\mathcal{L}_1$. It is well-known that for this nested model, for large sample size $n$, $W_n := -2(\mathcal{L}_0 - \mathcal{L}_1)$ follows approximately a chi-squared distribution with $2p - q$ degrees of freedom, where $2p$ and $q$ are the dimensions of $\Theta$ and $\Theta_0$, respectively. Once $W_n$ is computed, the test would reject the null hypothesis for large enough value of $W_n$.

We are interested in the following four null hypotheses, with the alternative being the full S1 distribution. The first two hypotheses are concerned with the parameters of the underlying small-circle (or sphere) $\mathcal{C}(\mu_0, \nu)$. The latter two validate the use of the S1 distribution over simpler distributions by testing whether the "small-circle feature" or the concentration around the mode is really there. Let $\theta = (\mu_0, \mu_1, \kappa_0, \kappa_1)$. The alternative hypothesis is then $H_1 : \theta \in \Theta = (\mathbb{S}^{p-1})^2 \times (\mathbb{R}_+)^2$.

1. $H_0$: $\mu_0 = \mu_0^*$, or $\theta \in \Theta_0 = \{\mu_0^*\} \times \mathbb{S}^{p-1} \times (\mathbb{R}_+)^2$. This is to test whether a predetermined axis $\mu_0^*$ of the small sphere is acceptable.

2. $H_0$: $\nu = 0$, or $\theta \in \Theta_0 \simeq \mathbb{S}^{p-1} \times \mathbb{S}^{p-2} \times (\mathbb{R}_+)^2$. ($A \simeq B$ means that $A$ and $B$ are diffeomorphic.) This is to test whether the underlying sphere has the radius 1, in which case it is a great sphere.

3. $H_0$: $\kappa_0 = 0$, or $\theta \in \Theta_0 \simeq \mathbb{S}^{p-1} \times \mathbb{R}_+$. Under $H_0$, the data follow a von Mises-Fisher distribution.

4. $H_0$: $\kappa_1 = 0$, or $\theta \in \Theta_0 \simeq \mathbb{S}^{p-1} \times \mathbb{R}_+$. Under $H_0$, the data follow a Bingham-Mardia distribution.

The null hypotheses 1, 2, and 4 above are also valid under the S2 distribution. For case 3, we can also test the null hypothesis $H_0 : \kappa_0 = 0$, but the null distribution is not a von Mises-Fisher distribution. The types of hypothesis are not limited to the above four examples. In particular,

testing the association between directional vectors is also valid and computationally feasible when using the MS2 distribution.

For each hypothesis, computing the test statistic $W_n$ requires to maximize the likelihood on $\Theta_0$ (or to compute $\mathcal{L}_0$). This is easily achieved by modifying the iterative algorithms in Section 3. For example, under the first null hypothesis (Case 1), where $\mu_0$ of the S1 distribution is fixed, one only needs to solve (23) in computation of $\mathcal{L}_0$; in Case 2, the fixed value of $\nu$ makes it enough to update the estimate $\hat{\mu}_0(\nu)$ once, thus solving (23) with fixed $\mu_0 = \hat{\mu}_0(\nu)$ provides the maximum likelihood estimates, restricted to $\Theta_0$. In the online supplementary material, we confirm that the test statistics $W_n$ computed from using our algorithms under the hypotheses listed above are empirically distributed as the chi-squared distributions. In Section 5.3 and in the online supplementary material, empirical powers of the proposed test procedures are reported for several important alternatives.

# 5 Numerical studies

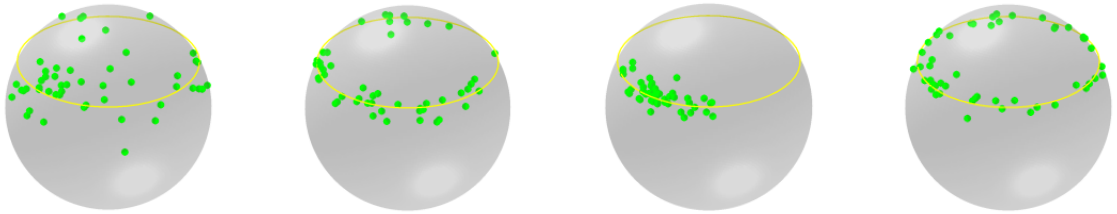
We demonstrate the performances of small-circle fitting in Section 5.1, an advantage of the MS2 by modeling the horizontal dependence in Section 5.2 and a testing procedure to prevent overfitting in Section 5.3.

## 5.1 Estimation of small-circles

The performance of our estimators in fitting the underlying small-spheres $\mathcal{C}(\mu_0, \nu)$ is numerically compared with those of the estimates obtained from assuming the Bingham-Mardia (BM, Bingham and Mardia, 1978) distribution and the least-square estimates of Schulz et al. (2015). The BM distribution is originally defined only for data on $\mathbb{S}^2$, but we use a natural extension given by a special case of the iMS1. In particular, "BM estimates" refer to the estimates of the iMS1 model with the restriction $\boldsymbol{\kappa}_1 = 0$. The estimates of Schulz et al. are obtained by minimizing the sum of squared angular distances from observations to $\mathcal{C}(\hat{\mu}_0, \hat{\nu})$, which will be referred to as a "least-squares (LS)" method.

We first considered four univariate S2 models to simulate data concentrated on a small circle. The directional parameters $(\mu_0, \mu_1)$ were set to satisfy $\nu = 0.5$. We used $(\kappa_0, \kappa_1) = (10, 1), (100, 1), (100, 10)$ to represent various data situations. Random samples from these three settings are shown in Fig. 4(a)–(c). We also considered the BM model as a special case of the S2 distributions (by setting $\kappa_1 = 0$); a sample from the BM distribution is shown in Fig. 4(d).

The small-circle estimation performances of the S1, S2, BM, and LS estimates are measured



(a) $\kappa_0 = 10, \kappa_1 = 1$ (b) $\kappa_0 = 100, \kappa_1 = 1$ (c) $\kappa_0 = 100, \kappa_1 = 10$ (d) $\kappa_0 = 100, \kappa_1 = 0$

Figure 4: Random samples from S2 on $\mathbb{S}^2$ used in our simulation settings

by an angular product error (in degrees), defined as

$$\mathcal{L}\left((\mu_0, \nu), (\hat{\mu}_0, \hat{\nu})\right) = \left(\text{Angle}(\hat{\mu}_0, \mu_0)^2 + \left(\frac{180}{\pi}(\cos^{-1}\hat{\nu} - \cos^{-1}\nu)\right)^2\right)^{1/2}.$$

Table 2 displays the means and standard deviations of $\mathcal{L}\left((\mu_0, \nu), (\hat{\mu}_0, \hat{\nu})\right)$ from 100 repetitions for each of the four methods, fitted to random samples of size 50 from each of the settings, labeled (a)–(d). The S2 estimates performed best for all non-trivial settings: (a), (b) and (c). Estimators under Settings (a) and (c) show higher errors. This is due to either the large vertical dispersion or the smaller horizontal dispersion as visualized in Fig. 4(a) and (c). Even when the sophisticated S1 and S2 models are not needed in Setting (d), the S1 and S2 estimators performed virtually the same as the BM estimator does.

Next, to show the performance of our multivariate models, we considered six bivariate MS2 models. The directional parameters $(\mu_0, \boldsymbol{\mu}_1)$ were set to satisfy $\boldsymbol{\nu} = (0.5, -.3)$, and the concentration parameters were chosen to mimic the concentrations of the univariate models, described above. For the first three cases (a)–(c), we set $(\kappa_{0j}, \kappa_{1j}, \lambda_{12}) = (10, 1, 0), (100, 1, 0), (100, 10, 0)$, for $j = 1, 2$, so that the models are indeed the iMS2. For the later three cases (d)–(f), we set $(\kappa_{0j}, \kappa_{1j}, \lambda_{12}) = (10, 2, 1.5), (100, 2, 1.5), (100, 20, 15)$, $j = 1, 2$, to make their vertical and horizontal dispersions be similar to the iMS2 counterparts. By setting non-zero $\lambda_{12}$, the random bivariate directions are positively associated. The small-circles estimation performance of the iMS1, iMS2, MS2, BM and LS estimates is measured by a simple multivariate extension of the angular product error. Table 3 collects the means and standard deviations of the angular product errors from 100 repetitions with the sample size $n = 50$. Overall, the three proposed models (iMS1, iMS2, and MS2) show better or competitive performances in the axis and radii estimation. In particular, when directions are clearly concentrated on small-circles and have horizontal dependence, i.e., in Settings (e) and (f), the MS2 estimates shows better performances than others.

| Method | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| S1 | 6.62(3.44) | 1.59(0.78) | 14.89(13.00) | **1.32**(0.56) |
| S2 | **6.06**(3.21) | **1.58**(0.76) | **14.57**(11.56) | 1.33(0.56) |
| BM | 9.54(9.80) | 1.66(0.81) | 16.59(13.57) | **1.32**(0.56) |
| LS | 6.48(3.37) | 1.61(0.75) | 14.68(11.54) | 1.33(0.55) |

Table 2: Small-circle estimation performances for univariate data on $\mathbb{S}^2$. Estimation accuracy is measured by the angular product error (in degrees).

| | Independent | | | Dependent | | |
|---|---|---|---|---|---|---|
| Method | (a) | (b) | (c) | (d) | (e) | (f) |
| iMS1 | 4.52(1.89) | 1.28(0.51) | **3.90(3.61)** | 7.01(2.88) | 1.58(0.76) | 4.71(5.39) |
| iMS2 | 4.45(1.71) | **1.27(0.51)** | 4.30(2.46) | **5.78(2.50)** | 1.58(0.75) | 4.60(2.69) |
| MS2 | **4.40(1.71)** | 1.28(0.51) | 4.26(2.45) | 5.90(2.46) | **1.57(0.75)** | **4.49(2.79)** |
| BM | 5.15(2.36) | 1.28(0.52) | 8.63(4.66) | 17.02(21.46) | 1.69(0.83) | 10.77(5.64) |
| LS | 4.48(1.86) | 1.28(0.53) | 4.30(2.35) | 6.81(3.20) | 1.59(0.71) | 4.62(2.82) |

Table 3: Small-circles estimation performances for bivariate data on $\mathbb{S}^2$.

We also checked a robustness of the estimators by simulating data from a signal-plus-noise model (neither S1 nor S2). The performances of small-circle fitting of the proposed methods are comparable to that of the least-square estimator. Relevant simulation results and a detailed discussion can be found in the online supplementary material.

### 5.2 Estimation of horizontal dependence

The ability of the MS2 to model the horizontal dependence is an important feature of the proposed distributions. Here, we empirically confirm that the MS2 estimates provide accurate measures of horizontal dependence, using the models (c) and (f) of the bivariate MS2, described above. For sample sizes $n = 50$ and 200, the concentration and association parameters were estimated under the assumption of MS2 and iMS2, and Table 4 summarizes the estimation accuracy for 100 repetitions. In all cases, the MS2 model provides precise estimations of the horizontal dispersion and dependence; as the sample size increases, the mean squared error decreases. For Case (c), the underlying model is exactly iMS2, so the iMS2 estimates have smaller mean squared errors than the MS2 estimates do. However, for Case (f), we notice that the iMS2 estimates of $\boldsymbol{\kappa}_1$ become inferior. In fact, when there exists a horizontal dependence, i.e., when $\lambda_{12} \neq 0$, the concentration parameter $\kappa_{1j}$ in the MS2 model must be interpreted differently from the $\kappa_{1j}$s in the iMS2 model. This is because the marginal distribution of $\phi_j$, $j = 1, 2$, in (16) is not a von Mises distribution (Shing et al., 2002; Mardia et al., 2008).

| | | (c) | | | (f) | | |
|---|---|---|---|---|---|---|---|
| $n$ | Method | $\kappa_{11}=10$ | $\kappa_{12}=10$ | $\lambda_{12}=0$ | $\kappa_{11}=20$ | $\kappa_{12}=20$ | $\lambda_{12}=15$ |
| 50 | iMS2 | 10.23(2.48) | 10.54(2.16) | | 11.73(2.19) | 11.19(2.14) | |
| | MS2 | 10.48(2.54) | 10.80(2.27) | −0.17(1.85) | 22.63(5.06) | 21.40(4.32) | 16.82(4.22) |
| 200 | iMS2 | 10.31(1.08) | 10.10(1.04) | | 11.00(1.05) | 11.09(1.04) | |
| | MS2 | 10.35(1.10) | 10.14(1.05) | −0.12(0.73) | 20.38(2.17) | 20.54(2.25) | 15.41(2.06) |

Table 4: Concentration and association parameters estimation performances for bivariate data on $\mathbb{S}^2$.

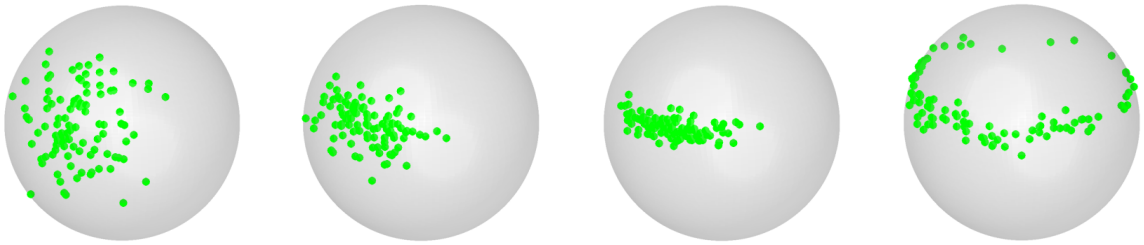
### 5.3 Small-circle estimation in an isotropic case

When the data do not exhibit a strong tendency of a small-circle feature, the S1 and S2 distributions may overfit the data. See for example a random sample from an isotropic distribution, shown in Fig. 5(a). To this data set, the S1 or S2 distribution nevertheless fits an unnecessary small-circle $\mathcal{C}(\hat{\mu}_0, \hat{\nu})$; a typical fitted small-circle is roughly matched to a contour of the vMF density, which was observed for 83% of the time, in a simulation of fitting the S1. Using the BM or the least-squares estimates results in a similar overfitting, where the extremely-small circles are erroneously fitted for 100% and 68% of the time, for the BM and least-squares respectively.

This problem of overfitting has been known for a while and discussed in the context of dimension reduction of directional data. In particular, Jung et al. (2011, 2012) investigated the overfitting phenomenon for the least-square estimates and proposed some ad-hoc methods for adjustment. To prevent the overfitting, we point out that the testing procedure in Section 4 for the detection of isotropic distribution (Case 3) works well. To confirm this, we evaluated the empirical power of the test at the significance level $\alpha = 0.05$ for several alternatives. The power increases sharply as the distributions become more anisotropic; under the alternative distributions depicted in Fig. 5(b)–(d), the empirical powers are respectively $\hat{\beta} = 0.435, 1$ and 1, evaluated from 200 repetitions.

## 6 Analysis of s-rep data

In this section, an application of the proposed distribution and test procedure to s-rep data is discussed.



(a) vMF ($\kappa = 10$) (b) $(\kappa_0, \kappa_1) = (20, 10)$ (c) $(\kappa_0, \kappa_1) = (100, 10)$ (d) $(\kappa_0, \kappa_1) = (100, 1)$

Figure 5: Degrees of the "small-circle feature." Shown are random samples from an isotropic distribution (case (a)) and the S2 distributions with increasing "small-circle concentrations" (cases (b)–(d)).

### 6.1 Modeling randomly-deformed ellipsoids via s-reps

The skeletal representations (s-reps) have been useful in mathematical modeling of human anatomical objects (Siddiqi and Pizer, 2008). Roughly, an s-rep model for a 3-dimensional object consists of locations of a skeletal mesh (inside of the object) and spoke vectors (directions and lengths), connecting the skeletal mesh with the boundary of the object; examples are shown in the top left panel of Fig. 6. When the object is "rotationally deformed", Schulz et al. (2015) have shown that the directional vectors of an s-rep model trace a set of concentric small-circles on $\mathbb{S}^2$, as shown in the top panels of the figure. Such rotational deformations (e.g., rotation, bending and twisting) of human anatomical objects have been observed in between and within shape variations of hippocampi and prostates (Joshi et al., 2002; Jung et al., 2011; Pizer et al., 2013). We demonstrate a use of the MS2 distribution in modeling (and fitting) a population of such objects via s-reps. Note that the sample space of an s-rep with $K$ spokes is $(\mathbb{S}^2)^K \times \mathbb{R}_+^K \times (\mathbb{R}^3)^K$ (for direction, length, and location). In this work, we choose to analyze the spoke directions in $(\mathbb{S}^2)^K$ only, leaving a full-on analysis, accommodating the lengths and locations, as a future work.

### 6.2 Data preparation

For our purpose of validating the use of the proposed distributions, we use an s-rep data set, fitted from 30 deformed ellipsoids; three samples from this data set are shown in Fig. 6. This data set was previously used in Schulz et al. (2015) as a simple experimental representation of real human organs. The data set was generated by "physically bending" a template ellipsoid about an axis $\mu_0^* = (0, 1, 0)$ by random angles drawn from a normal distribution with standard deviation 0.4 (radians). Each deformed ellipsoid is then recorded as a 3-dimensional binary image. To mimic the procedure of fitting s-reps from, for example, medical resonance imaging of

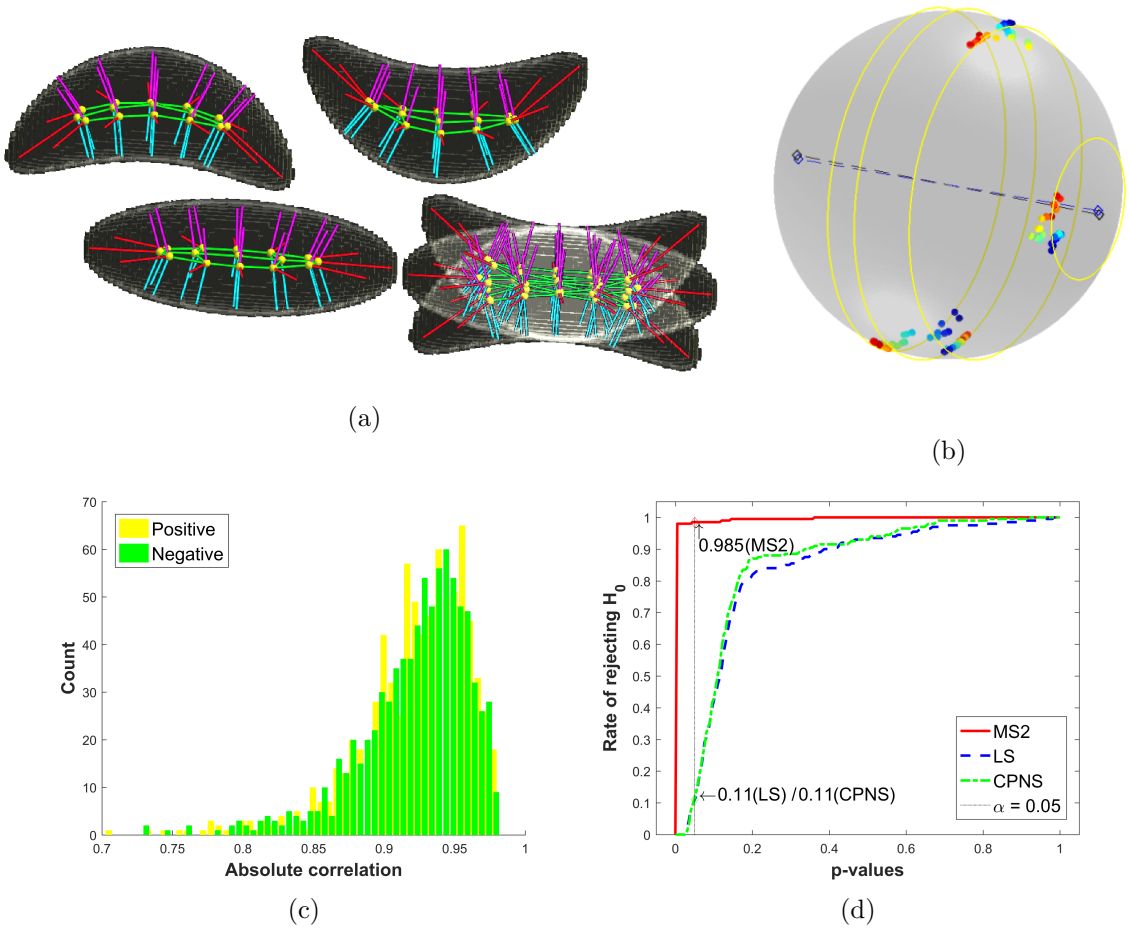



Figure 6: Figures in (a): S-reps modeling of randomly-bent ellipsoids overlaid with corresponding binary images. (b) Directions-circles plot: Graphical display of MS2 parameter estimates (small-circles (yellow), $\hat{\mu}_0$ (blue dashed axis) compared with $\mu_0^*$ (black dashed)) overlaid with the data where different colors correspond to different observations. (c) Histogram of estimated "horizontal" correlation coefficients. (d) Empirical p-values from horizontal-dependence tests. See text for details.

a real patient, s-reps with 74 spokes were fitted to these binary images. (See Pizer et al. (2013) for details of the s-rep fitting.) As a preprocessing, we chose $K = 58$ spoke vectors, excluding the vectors with very small total variation.

### 6.3 Inference on the bending axis

Assuming the iMS1 or MS2 distributions, we obtained estimates of the axis $\hat{\mu}_0^{(\text{iMS1})} = (0.007, 1.000, -0.008)$ and $\hat{\mu}_0^{(\text{MS2})} = (0.006, 1.000, 0.006)$ (rounded to three decimal digits). These estimates are virtually the same, only 0.6 degrees away from the ground truth $\mu_0^*$. Estimates of the concentric small-circles $\mathcal{C}(\hat{\mu}_0^{(\text{MS2})}, \hat{\nu}_j)$ for four choices of $j$ (the spoke index) are also shown in the top right

panel of Fig. 6 in which $\hat{\mu}_0^{(\mathrm{MS2})}$ and $\mu_0^*$ are also overlaid. The test procedure discussed in Section 4 (Case 1) was applied to test $H_0 : \mu_0 = \mu_0^*$ under the iMS1 model. With a test statistics $W_n = 2.69$, compared to 5.99 (the 95th percentile of $\chi^2_2$), and the corresponding p-value of 0.26, we confirm that the s-rep spokes are rotated about the bending axis $\mu_0^*$.

### 6.4 Inference on horizontal dependence

An advantage of modeling the s-rep spoke directions by the MS2 distribution is the ability of perceiving and modeling the horizontal dependence among directions. As an exploratory step, we have collected the estimated correlation coefficients, computed from the approximate precision matrix $\widehat{\Sigma}^{-1}$, whose elements are $\hat{\boldsymbol{\kappa}}_1$ and $\widehat{\boldsymbol{\Lambda}}$; see (17). A histogram of $K(K-1)/2$ estimated correlation coefficients is plotted in the bottom left panel of Fig. 6. As expected, pairs of spoke vectors from the same side (e.g. two spoke vectors in the "left side" of the ellipsoids in Fig. 6) exhibit strong positive correlations, while those from the opposite sides exhibit strong negative correlations. The horizontal dependence is in fact apparent by the way data were generated (simultaneously bending all the spoke directions).

For large enough sample sizes, we could use the likelihood-ratio test procedure discussed in Section 4 for testing $H_0 : \boldsymbol{\Lambda} = \mathbf{0}$ using the MS2 distribution (16). This is infeasible, due to our small sample size, $n = 30$, and the large number of parameters tested, 1653 ($= K(K-1)/2$). Coping with this high-dimension, low-sample-size situation is beyond the scope of current paper, and we resort to choose only two or three spoke directions to test the dependence, but to repeat the testing for arbitrarily many combinations of total $K = 58$ spokes. For the bivariate case, applying the likelihood-ratio test procedure produces a p-value for each pair of spokes. In the bottom right panel of Fig. 6, we have plotted the empirical distribution of these p-values. It can be seen that, at the significance level 0.05, the "MS2" test of dependence is indeed powerful, with a rejection rate of 97%.

To provide some context to this rate, the MS2 test was compared with other natural choices of tests. We applied two methods that were previously used for s-rep data: the least-square (concentric) small-circle fitting method of Schulz et al. (2015) and the composite principal nested spheres (CPNS), discussed in Pizer et al. (2013). Briefly, the CPNS fits a least-square small-circle to each marginal directions on $\mathbb{S}^2$, the observations are then each transformed to spherical coordinates (with an understanding that the axis of the fitted small-circle is the north pole). For the purpose of testing "horizontal associations", we only keep the longitudinal coordinates, which then form a bivariate angular data set. Assuming normality, we simply test whether correlation coefficient is zero. We refer to this test procedure by a "CPNS" test. An "LS" test can be performed similarly, with the individual small-circle fits replaced by the fitted concentric

small-circles. These two tests were also conducted for the same combinations of spoke directions, and the empirical distributions of respective p-values are also plotted in Fig. 6. These alternative tests appear to be too conservative, with rejection rates 11% for the LS test, and 13.5% for the CPNS test (at level 0.05). Heuristically, the higher power of the MS2 tests is due to the superior fitting of the MS2 distribution. In particular, the "horizontal angles" predicted from the MS2 tend to be linearly associated, while those from the least-squares fit tend to be arbitrary. We refer to the online supplementary material for more simulation results. All in all, using the MS2 distribution shows a clear advantage in modeling and testing the horizontal dependence of multivariate directions.

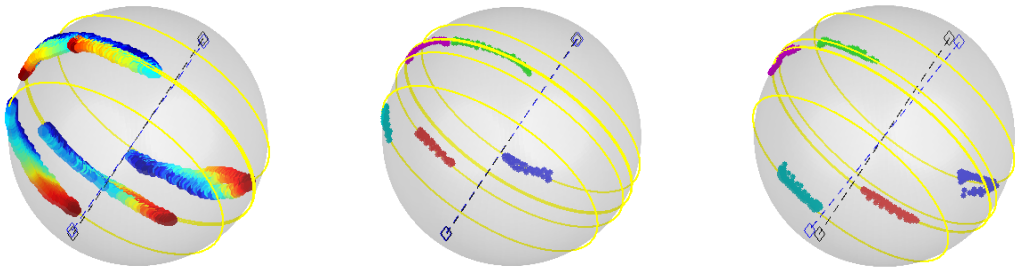
# 7 Human knee gait analysis

In biomechanical gait analysis, accurately modeling human knee motion during normal walking has a potential to differentiate diseased subjects from normal subjects. In particular, the axis of bending (of the lower leg toward the upper leg) is believed to be a key feature in the discrimination among the diseased and normal subjects (Pierrynowski et al., 2010). As a step towards the development of statistical tests for "two-group axes difference," in this section we contemplate using the proposed distribution families in modeling the bending motion of the knee.

The raw data set we use is obtained from healthy volunteer and is a time series of coordinates of markers planted at the volunteer's leg, recorded for 16 gait cycles. For each time point, the direction vectors on $(\mathbb{S}^2)^5$ were computed to be the unit tangential vectors from a reference marker to others, as done in Schulz et al. (2015). We treat the data set as if the observations were independent, leaving a proper time-series analysis of directions as a future work. Nonetheless, these direction vectors are *horizontally dependent* to each other (as evidenced in Fig. 7), which suggests that we may fit the MS2 distribution.

The first panel of Fig. 7 illustrates the result of MS2 fit to the all data points. There, we superimposed the fitted concentric circles to the observed direction vectors, together with a hypothesized dominant bending axis $\mu_0^* = (0, 1, 0)^\top$, the left-right axis of the subject. The MS2 model seemed to fit well with high estimated horizontal correlation coefficients. We, however, identified a strong evidence against using the MS2. Specifically, as shown in Fig. 7 some direction vectors exhibit higher variations (e.g. the one corresponding to the north-most circle) for a subset of time points.

In fact, the data are better viewed as a mixture of many periods of the gait cycle. In particular, we focused on the "swing" and "stance" periods, and separately analyzed subsampled data from each period. For the swing period, the MS2 model fits well (see Fig. 7(b)), and the axis of bending is estimated as $\hat{\mu}_0^{(\mathrm{SW})} = (0.013, 1.000, 0.005)^\top$, only 0.8 degrees away from the



(a) All data points (b) Swing period (c) Stance period

Figure 7: Knee gait analysis: Observed direction vectors overlaid with the hypothesized (black dashed) and estimated (blue dashed) axes as well as the MS2-fitted small circles. In (a), colors code time indices. In (b) and (c), different colors represent different $\mathbb{S}^2$.

hypothesized axis, $\mu_0^*$. Our likelihood ratio test procedure does not reject the null hypothesis $H_0 : \mu_0 = \mu_0^*$, with p-value 0.16. On the other hand, the MS2 model does not fit well for the stance period, again due to some irregular patterns in the data. Excluding the highly-irregular direction, the MS2 fits well, and our test procedure confirms that the axis of bending for this period differs from $\mu_0^*$, with p-value less than $10^{-5}$. The estimated axis for the stance period is $\hat{\mu}_0^{(\mathrm{st})} = (0.11, 0.994, 0.006)^\top$.

While the MS2 distribution was useful in making inference on the bending axis of partial knee motions, future work for this type of data lies in the development of a two-sample axis difference test and a functional-and-directional data modeling.

# A Appendix

We provide a technical lemma, referenced in Section 3.1, and proofs of Propositions 1 and 2.

**Lemma 3.** *If $X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$, then $E(X)$ is a linear combination of $\mu_0$ and $\mu_1$.*

*Proof of Lemma 3.* Suppose that for some $a, b, c \in \mathbb{R}$, $v \in S^{p-1}$, $E(X) = a\mu_0 + b\mu_1 + cv$. Then choose a $B \in O(p)$ such that $B\mu_0 = \mu_0$, $B\mu_1 = \mu_1$ but $Bv \neq v$. By Proposition 1(i), $BX \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$. Thus $E(X) = E(BX)$, which in turn leads to $a\mu_0+b\mu_1+cv = a\mu_0+b\mu_1+cBv$, which is true only if $c = 0$. This gives the result. □

*Proof of Proposition 1.* Assertions (i) and (iii) are trivial. (ii): Assume that $X \sim BX$. With

the representation (10) this is equivalent with

$$(B\gamma)^\top x = \gamma^\top x \text{ and } \left((B\mu_0)^\top x\right)^2 = (\mu_0^\top x)^2 \text{ for all } x \in \mathbb{S}^{p-1}\,.$$

In consequence

$$2\nu\kappa_0 B\mu_0 + \kappa_1 B\mu_1 = B\gamma = \gamma = 2\nu\kappa_0\mu_0 + \kappa_1\mu_1 \text{ and } B\mu_0 = \pm\mu_0\,.$$

Thus the case of $B\mu_0 = \mu_0$ yields at once $B\mu_1 = \mu_1$. In contrast, the case $B\mu_0 = -\mu_0$ yields that $\|B\mu_1\|^2 = \|4\nu\frac{\kappa_0}{\kappa_1}\mu_0 + \mu_1\|^2 = 1 + 8\nu^2\kappa_0/\kappa_1(1 + 2\kappa_0/\kappa_1) > 1$, which cannot be if $B$ is orthogonal.

Further, if $Y \sim BY$, since the exponent is again the sum of an even function in $x$ and an odd function, we have again that $B\mu_0 = \pm\mu_0$, yielding $\|P_{\mu_0}x\| = \|P_{\mu_0}Bx\|$, due to orthogonality, $B$ preserves the space orthognal to $\mu_0$, as well as

$$2\nu\kappa_0\mu_0^\top x + \frac{\kappa_1}{\|P_{\mu_0}\mu_1\|\,\|P_{\mu_0}x\|}\mu_1^\top(I_p - \mu_0\mu_0^\top)x = 2\nu\kappa_0(B\mu_0)^\top x + \frac{\kappa_1}{\|P_{\mu_0}\mu_1\|\,\|P_{\mu_0}x\|}\mu_1^\top(I_p - \mu_0\mu_0^\top)B^\top x\,.$$

As before, $B\mu_0 = \mu_0$ yields that $B\mu_1 = \mu_1$ and $B\mu_0 = -\mu_0$ would give that

$$\left((B\mu_1)^\top - \mu_1^\top\right)x = 2\nu\left(\frac{2\kappa_0}{\kappa_1}\sqrt{1-\nu^2}\sqrt{1-(\mu_0^\top x)^2} - 1\right)\mu_0^\top x \text{ for all } x \in \mathbb{S}^{p-1}\,.$$

In particular, this would yield that $B\mu_1 - \mu_1$ is a multiple of $\mu_0$, the factor, however, is not constant in $x$, a contradiction. □

*Proof of Proposition 2.* For a given $h > 0$, let $\gamma = 2\kappa_0\nu\mu_0 + \kappa_1\mu_1$ and $A_h = \kappa_0\mu_0\mu_0^\top + hI_p$. Then it is easy to see that the S1 density (5) can be expressed as the Fisher-Bingham form (10):

$$f_{\mathrm{S1}}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \alpha(\gamma, A_h)\exp\{\gamma^\top x - x^\top A_h x\}$$

where $\alpha(\gamma, A_h)$ satisfies

$$a(\kappa_0, \kappa_1, \nu) = \alpha(\gamma, A_h)\exp\{-\kappa_0\nu^2 + h\}. \tag{32}$$

In the purpose of evaluating the value of $a(\kappa_0, \kappa_1, \nu)$, or equivalently $\alpha(\gamma, A_h)$ for the given value of $h$, one can assume without losing generality that $\mu_0 = (1, 0, \ldots, 0)^\top$ and $\mu_1 = (\nu, \sqrt{1-\nu^2}, 0, \ldots, 0)$, so that $\gamma = (\nu(2\kappa_0 + \kappa_1), \kappa_1\sqrt{1-\nu^2}, 0, \ldots, 0)^\top$, and the vector of diagonal values of $A_h$ are $\lambda := (2(\kappa_0 + h), h, \ldots, h)$. The $j$th element of $\xi$, in the statement of proposition, is then given by $\mu_j := \gamma_j/2\lambda_j$. With these notations, Proposition 1 of Kume and Wood (2005) gives

$$\alpha(\Psi, \mu) = 2\pi^{p/2}|A_h|^{-1/2}g(1)\exp\{\xi^\top A_h\xi\}.$$

Hence, by (32), we have (18). □

## References


Albers, C., Critchley, F., and Gower, J. (2011), "Quadratic minimisation problems in statistics," *Journal of Multivariate Analysis*, 102, 698–713.

Banerjee, A., Dhilon, I. S., Ghosh, J., and Sra, S. (2005), "Clustering on the unit hypersphere using von Mises-Fisher distributions," *Journal of Machine Learning Research*, 6, 1345–1382.

Bingham, C. and Mardia, K. V. (1978), "A small circle on the sphere distribution," *Biometrika*, 65, 379–389.

Browne, M. W. (1967), "On oblique Procrustes rotation," *Psychometrika*, 32, 125–132.

Cootes, T. F., Taylor, C., Cooper, D., and Graham, J. (1992), "Training Models of Shape From Sets of Examples," in *Proceedings of British Machine Vision Conference*, eds. Hogg, D. and Boyle, R., Berlin: Springer-Verlag, pp. 9–18.

Dryden, I. and Mardia, K. V. (1998), *Statistical Shape Analysis*, Chichester: Wiley.

Gray, N. H., Geiser, P. A., and Geiser, J. R. (1980), "On the least-squares fit of small and great circles to spherically projected orientation data," *Journal of the International Association for Mathematical Geology*, 12, 173–184.

Hoff, P. D. (2009), "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data," *Journal of Computational and Graphical Statistics*, 18, 438–456.

Joshi, S., Pizer, S., Fletcher, P. T., Yushkevich, P., Thall, A., and Marron, J. S. (2002), "Multi-scale deformable model segmentation and statistical shape analysis using medial descriptions," *IEEE transactions on medical imaging*, 21, 538–50.

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568.

Jung, S., Foskey, M., and Marron, J. S. (2011), "Principal arc analysis on direct product manifolds," *The Annals of Applied Statistics*, 5, 578–603.

Kent, J. T. (1982), "The Fisher-Bingham distribution on the sphere," *Journal of the Royal Statistical Society. Series B (Methodological)*, 71–80.

Kume, A., Preston, S., and Wood, A. T. (2013), "Saddlepoint approximations for the normalizing constant of Fisher–Bingham distributions on products of spheres and Stiefel manifolds," *Biometrika*, 100, 971–984.

Kume, A. and Wood, A. T. A. (2005), "Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants," *Biometrika*, 92, 465–476.

Kurtek, S., Ding, Z., Klassen, E., and Srivastava, A. (2011), "Parameterization-Invariant Shape Statistics and Probabilistic Classification of Anatomical Surfaces," in *Information Processing in Medical Imaging*, Berlin: Springer, vol. 6801, pp. 147–158.

Mardia, K. (1975), "Statistics of directional data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 37, 349–393.

Mardia, K. V. and Gadsden, R. J. (1977), "A Small Circle of Best Fit for Spherical Data and Areas of Vulcanism," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26, 238–245.

Mardia, K. V., Hughes, G., Taylor, C. C., and Singh, H. (2008), "A multivariate von Mises distribution with applications to bioinformatics," *Canadian Journal of Statistics*, 36, 99–109.

Mardia, K. V. and Jupp, P. E. (2000), *Directional Statistics*, vol. 28 of *Wiley series in probability and statistics*, Wiley.

Pierrynowski, M., Costigan, P., Maly, M., and Kim, P. (2010), "Patients With Osteoarthritic Knees Have Shorter Orientation and Tangent Indicatrices During Gait," *Clinical Biomechanics*, 25, 237–141.

Pizer, S. M., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., and Marron, J. S. (2013), "Nested Sphere Statistics of Skeletal Models," in *Innovations for Shape Analysis: Models and Algorithms*, eds. Breu, M., Bruckstein, A., and Maragos, P., New York: Springer, pp. 93–115.

Rivest, L. P. (1999), "Some Linear Model Techniques for Analyzing Small-Circle Spherical Data," *Canadian Journal of Statistics*, 27, 623–638.

Schulz, J., Jung, S., Huckemann, S., Pierrynowski, M., Marron, J. S., and Pizer, S. M. (2015), "Analysis of Rotational Deformations from Directional Data," *Journal of Computational and Graphical Statistics*, 24, 539–560.

Shing, H., Hnizdo, V., and Demchuk, E. (2002), "Probabilistic Model for Two Dependent Circular Variables," *Biometrika*, 89, 719–723.

Siddiqi, K. and Pizer, S. (2008), *Medial Representation: Mathematics, Algorithms and Applications*, Springer.