Electronic Journal of Statistics
Vol. 15 (2021) 984–1033
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1807

Geodesic projection of the von Mises–Fisher distribution for projection pursuit of directional data*

Sungkyu Jung

Department of Statistics, Seoul National University
Seoul 08826, South Korea
e-mail: sungkyu@snu.ac.kr

Abstract: We investigate geodesic projections of von Mises–Fisher (vMF) distributed directional data. The vMF distribution for random directions on the $(p-1)$-dimensional unit hypersphere $\mathbb{S}^{p-1} \subset \mathbb{R}^p$ plays the role of multivariate normal distribution in directional statistics. For one-dimensional circle $\mathbb{S}^1$, the vMF distribution is called von Mises (vM) distribution. Projections onto geodesics are one of main ingredients of modeling and exploring directional data. We show that the projection of vMF distributed random directions onto any geodesic is approximately vM-distributed, albeit not exactly the same. In particular, the distribution of the geodesic-projected score is an infinite scale mixture of vM distributions. Approximations by vM distributions are given along various asymptotic scenarios including large and small concentrations ($\kappa \to \infty, \kappa \to 0$), high-dimensions ($p \to \infty$), and two important cases of double-asymptotics ($p, \kappa \to \infty$, $\kappa/p \to c$ or $\kappa/\sqrt{p} \to \lambda$), to support our claim: geodesic projections of the vMF are approximately vM. As one of potential applications of the result, we contemplate a projection pursuit exploration of high-dimensional directional data. We show that in a high dimensional model almost all geodesic-projections of directional data are nearly vM, thus measures of non-vM-ness are a viable candidate for projection index.

MSC2020 subject classifications: Primary 62H11; secondary 62E20, 62R30.
Keywords and phrases: Bessel function, Struve function, projection index, von Mises distribution.

Received May 2020.

Contents

*This work was supported by the National Research Foundation of Korea grant funded by the Korea government (MSIT) (No. 2019R1A2C2002256).

1. Introduction

The von Mises–Fisher (vMF) distribution, sometimes referred to as the Fisher–von Mises–Langevin distribution, plays a central role in modeling and inference for the directional data (Mardia and Jupp, 2000; Ley and Verdebout, 2017). While the multivariate normal distribution is not defined on the sample space $\mathbb{S}^{p-1} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{x}'\mathbf{x} = 1\}$ of directions, the vMF distribution is the closest notion to the normal distribution. For a random direction $\mathbf{x} \in \mathbb{S}^{p-1}$, we write $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$ if its density function is $f_{\mathrm{vMF}}(\mathbf{x}) = c_p(\kappa)\exp(\kappa\mathbf{x}'\boldsymbol{\mu})$ with respect to the surface area measure of $\mathbb{S}^{p-1}$, for the location parameter $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$ and the concentration parameter $\kappa \geq 0$. The normalizing constant is

$$c_p(\kappa) = \frac{\kappa^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(\kappa)}, \tag{1.1}$$

where I_ν is the modified Bessel function of the first kind of order ν. (We use the convention $0/0 = 1$.) On $\mathbb{S}^1$, the vMF distribution becomes the von Mises (vM) distribution.

We consider the one-dimensional projection of directional data onto geodesics, a notion closest to the orthogonal projection onto a vector in Euclidean space. Projections onto geodesics and the evaluation of the projection scores and residuals are fundamental ingredients of modern directional statistics and, in general, statistics on manifolds; applications include dimension reduction (Fletcher et al., 2004; Jung, Dryden and Marron, 2012), regression (Fletcher, 2013; Cornea et al., 2017), classification (Pizer and Marron, 2017) and developments of parametric models (Schulz et al., 2015; Kim et al., 2019). Here, the projection onto a geodesic γ is defined intrinsically. With a distance function $\rho : \mathbb{S}^{p-1} \times \mathbb{S}^{p-1} \to [0, \pi]$, the projection of $\mathbf{x} \in \mathbb{S}^{p-1}$ onto γ is the point on γ closest to $\mathbf{x}$, and the projection score $S_\gamma(\mathbf{x}) \in (-\pi, \pi]$ is given by a parameterization of γ. See Section 2 for definitions of the geodesic, the metric, and the geodesic projection.

In Euclidean space, the projection $\mathbf{u}'\mathbf{z}$ of a Gaussian random vector $\mathbf{z} \in \mathbb{R}^p$ is also Gaussian for any $\mathbf{u} \in \mathbb{S}^{p-1}$, and even almost all projections of a non-Gaussian $\mathbf{z}$ are also Gaussian, asymptotically; see Diaconis and Freedman (1984); Bickel, Kur and Nadler (2018). On the other hand, it is yet unknown whether the geodesic projection $S_\gamma(\mathbf{x})$ of a vMF-distributed random direction $\mathbf{x} \in \mathbb{S}^{p-1}$ is also vM.

In this paper, we show that the answer to the above fundamental question is indeed negative. In particular, we show that for $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$ and for any geodesic γ, the distribution of $S_\gamma(\mathbf{x})$ is an infinite scale mixture of vM distributions with concentration parameters ranging from 0 to κ. Nevertheless, the density of $S_\gamma(\mathbf{x})$ is *nearly* vM-distributed, in the sense that it is close to the vM family to human eyes. The difference is not statistically significant for moderately large sample sizes.

The distribution of $S_\gamma(\mathbf{x})$ is denoted by $\mathrm{PvM}(p, \kappa, \delta)$, where $\delta = \rho(\gamma, \boldsymbol{\mu}) \geq 0$ measures how much γ deviates from the mode of $\mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$. In a special case where the geodesic γ passes through the mode, $\delta = 0$ and we say the geodesic is canonical with respect to $\boldsymbol{\mu}$. For canonical geodesics γ, the density function of $S_\gamma(\mathbf{x})$ is expressed using I_ν and M_ν, the modified Struve function of the second kind of order ν, to facilitate approximations by vM densities. We refer to Appendix A for definitions and several useful properties of I_ν and M_ν.

We show that $\mathrm{PvM}(p, \kappa, \delta)$ is indeed well approximated by vM distributions in various asymptotic scenarios, including large and small concentrations ($\kappa \to \infty$ and $\kappa \to 0$), high dimensions ($p \to \infty$), and a double-asymptotic direction $p \to \infty, \kappa \to \infty, \kappa/p = c \in (0, \infty)$. Various asymptotic expansions of I_ν and M_ν are used in the approximations. For the special asymptotic regime of $p \to \infty, \kappa \to \infty, \kappa/\sqrt{p} = \lambda \in (0, \infty)$, we provide a high dimensional approximation of the distribution by the projected normal distribution (Mardia and Jupp, 2000, p. 178). The projected normal distribution is again approximated by the vM distribution.

This work is developed in the course of formulating projection pursuit for directional data. Projection pursuit (Friedman and Tukey, 1974) is a well-

developed technique for exploration of multivariate data, aiming at finding low-dimensional "interesting" projections of high-dimensional data. See Friedman and Stuetzle (1981); Huber (1985); Friedman (1987); Jee (2009); Alashwali and Kent (2016); Bickel, Kur and Nadler (2018) and references therein for details on projection pursuit. While there have been various proposals for the index of interestingness, most of those measure departures from normality. For example, Bickel, Kur and Nadler (2018) compared the empirical distribution $\hat{F}_{\mathbf{u}'\mathbf{x}}$ directly with the normal distribution function, and Huber (1985) discussed the negative entropy $\int f \log f$ as an index of interestingness, which is uniquely minimized when the density f of the projection score is the normal density.

For directional data on $\mathbb{S}^{p-1}$ and their projections onto a geodesic (intrinsically a circle $\mathbb{S}^1$), the vMF and vM distribution play the role of the normal distribution. In particular, the vMF density minimizes the negative entropy among all choices of density on $\mathbb{S}^{p-1}$ (with fixed location and scale) as shown in Mardia (1975). Therefore, we propose to use a measure of discrepancy from the vM distribution as a one-dimensional projection pursuit index for directional data. To examine the potential of such an index, we show for a location mixture of vMF distributions the projection score is approximately vM-distributed in high dimensions. In two low-dimensional examples, using the p-value from a goodness-of-fit test, compared with the vM family, as the measure of uninterestingness, we demonstrate that clusters and outliers stand out in the projection score with small p-values.

The rest of article is organized as follows. In Section 2, we formally define the geodesic projection and projection score. In Section 3, the density of the geodesic-projected scores of vMF-distributed direction, $\mathrm{PvM}(p, \kappa, \delta)$, is evaluated. The density of PvM is compared to the Jones–Pewsey distribution family (Jones and Pewsey, 2005), which includes the vM distribution. In Section 4, we provide various asymptotic approximations of $\mathrm{PvM}(p, \kappa, \delta)$ by the vM, normal, Cardioid and the projected normal distributions. A numerical comparison of the approximations is deferred to Appendix C. Section 5 demonstrates the potential of using non-vM-ness as a projection index, towards a development of projection pursuit for directional data. We show that almost all projections are nearly vM in high dimensions. Issues in developing projection pursuit for directional data are pointed out in Section 6. Special functions and their properties are listed in Appendix A, and nearly all proofs are contained in Appendix B.

2. Geodesic projections on the unit hypersphere

Geodesics on the unit hypersphere $\mathbb{S}^{p-1}$ are the notion closest to straight lines in Euclidean space. A geodesic is a constant-velocity curve $\gamma : \mathbb{R} \to \mathbb{S}^{p-1}$ that locally minimizes the length of the curve among curves whose endpoints are fixed. On $\mathbb{S}^{p-1}$, geodesic paths coincide with great circles, and a geodesic γ can be parametrized with $\mathbf{q} \in \mathbb{S}^{p-1}$ and $\mathbf{v} \in T_{\mathbf{q}}$, where $T_{\mathbf{q}} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{x}'\mathbf{q} = 0\}$ is the tangent space of $\mathbb{S}^{p-1}$ at $\mathbf{q}$. A vector $\mathbf{u} \in T_{\mathbf{q}}$ can be mapped to $\mathbb{S}^{p-1}$ by the Exponential map, defined as $\mathrm{Exp}_{\mathbf{q}}(\mathbf{u}) = \sin(\|\mathbf{u}\|)\mathbf{u}/\|\mathbf{u}\| + \cos(\|\mathbf{u}\|)\mathbf{q}$. The

geodesic parameterized by $(\mathbf{q}, \mathbf{v})$ is

$$\gamma(t) = \gamma(\mathbf{q}, \mathbf{v})(t) = \mathrm{Exp}_{\mathbf{q}}(t\mathbf{v})$$

for $t \in \mathbb{R}$. Throughout, we require $\|\mathbf{v}\| = 1$ so that $\mathbf{v} \in \mathbb{S}^{p-1}$ and $\mathbf{v}'\mathbf{q} = 0$. With the restriction $\|\mathbf{v}\| = 1$, the image of $\mathbb{R}$ under γ, $\gamma(\mathbb{R})$, is equal to the image of $I := (-\pi, \pi]$ under γ. That is, $\gamma(I) = \gamma(\mathbb{R}) = \{\mathrm{Exp}_{\mathbf{q}}(t\mathbf{v}) : t \in \mathbb{R}\}$.

The geodesic distance function $\rho : \mathbb{S}^{p-1} \times \mathbb{S}^{p-1} \to [0, \pi]$ between $\mathbf{x}, \mathbf{y} \in \mathbb{S}^{p-1}$ is defined as the length of the shortest path on $\mathbb{S}^{p-1}$ between the two points, and is the arc length formed by the geodesic segment joining $\mathbf{x}$ and $\mathbf{y}$: $\rho(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\mathbf{x}'\mathbf{y})$. For a $\mathbf{w} \in T_{\mathbf{x}}$ satisfying $\mathbf{y} = \mathrm{Exp}_{\mathbf{x}}(\mathbf{w})$ and $\|\mathbf{w}\| \le \pi$, $\rho(\mathbf{x}, \mathbf{y}) = \|\mathbf{w}\|$. If such $\mathbf{w}$ is unique, then the shortest path between $\mathbf{x}$ and $\mathbf{y}$ is given by the geodesic joining $\mathbf{x}$ and $\mathbf{y}$, denoted by $\Gamma(\mathbf{x} \to \mathbf{y})$.

Definition 1 (Geodesic projection and score)**.**

(i) The projection of $\mathbf{x} \in \mathbb{S}^{p-1}$ onto the geodesic $\gamma(\mathbb{R})$ is the point on $\gamma(\mathbb{R})$ closest to $\mathbf{x}$ (in terms of the geodesic distance), and is denoted by $P_\gamma(\mathbf{x}) := \mathrm{argmin}_{\mathbf{p} \in \gamma(\mathbb{R})} \rho(\mathbf{p}, \mathbf{x})$.

(ii) The projection score $S_\gamma(\mathbf{x})$ of $\mathbf{x}$ onto $\gamma(\mathbb{R})$, defined with the parameterization $\gamma = \gamma(\mathbf{q}, \mathbf{v})$, is $t \in (-\pi, \pi]$ satisfying $P_\gamma(\mathbf{x}) = \mathrm{Exp}_{\mathbf{q}}(t\mathbf{v})$.

The geodesic projection and scores for an $\mathbb{S}^2$-valued sample are illustrated in Fig. 1. For any two points $\mathbf{x}_1, \mathbf{x}_2$, the geodesics $\Gamma(\mathbf{x}_1 \to P_\gamma(\mathbf{x}_1))$ and $\Gamma(\mathbf{x}_2 \to P_\gamma(\mathbf{x}_2))$ are not in general parallel.

The projection $P_\gamma(\mathbf{x})$ and the score $S_\gamma(\mathbf{x})$ are invariant and equivariant, respectively, to reparameterizations of $\gamma(\mathbf{q}, \mathbf{v})$. Specifically, let $\mathbf{q}_0$ and $\mathbf{v}_0$ be any orthogonal unit vectors satisfying $\mathrm{span}(\mathbf{q}, \mathbf{v}) = \mathrm{span}(\mathbf{q}_0, \mathbf{v}_0)$. Let $\theta = \cos^{-1}(\mathbf{q}'\mathbf{q}_0)$ and $r = \det([\mathbf{q}, \mathbf{v}]'[\mathbf{q}_0, \mathbf{v}_0]) \in \{1, -1\}$ (representing the rotation and reflection parts of the reparameterization). Then, the reparameterized geodesic $\gamma_0 = \gamma(\mathbf{q}_0, \mathbf{v}_0)$ satisfies $\gamma_0(\mathbb{R}) = \gamma(\mathbb{R})$, and we have $P_\gamma(\mathbf{x}) = P_{\gamma_0}(\mathbf{x})$ and $S_\gamma(\mathbf{x}) = rS_{\gamma_0}(\mathbf{x}) + \theta$, for any $\mathbf{x} \in \mathbb{S}^{p-1}$.

The next lemma is on the uniqueness and computation of $P_\gamma(\mathbf{x})$ and $S_\gamma(\mathbf{x})$.

Lemma 2.1. *Given $\mathbf{x} \in \mathbb{S}^{p-1}$ and a geodesic $\gamma(\mathbf{q}, \mathbf{v})$, if either $\mathbf{x}'\mathbf{q}$ or $\mathbf{x}'\mathbf{v}$ is nonzero, then $P_\gamma(\mathbf{x})$ is uniquely given by $P_\gamma(\mathbf{x}) = \mathrm{Exp}_{\mathbf{q}}(S_\gamma(\mathbf{x})\mathbf{v})$, where*

$$S_\gamma(\mathbf{x}) = \arg\min_{t \in \mathbb{R}} \rho(\mathrm{Exp}_{\mathbf{q}}(t\mathbf{v}), \mathbf{x}) = \mathrm{atan2}(\mathbf{x}'\mathbf{v}, \mathbf{x}'\mathbf{q}).$$

Otherwise, $\mathbf{x}'\mathbf{q} = \mathbf{x}'\mathbf{v} = 0$ and $P_\gamma(\mathbf{x})$ is everywhere on $\gamma(\mathbb{R})$.

The function $\mathrm{atan2}(y, x)$ is the two-argument inverse tangent which takes the signs of x and y into account to determine in which quadrant (x, y) lies. In other words, $S_\gamma(\mathbf{x})$ is the polar angle of the polar coordinates for $(\mathbf{x}'\mathbf{q}, \mathbf{x}'\mathbf{v})$ in Cartesian coordinates.

3. Geodesic projections of vMF random directions

For $\mathbf{x} \in \mathbb{S}^{p-1}$, suppose $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$. We wish to find the distribution of the projection score $S_\gamma(\mathbf{x})$ on any given γ. While vMF plays the role of normal

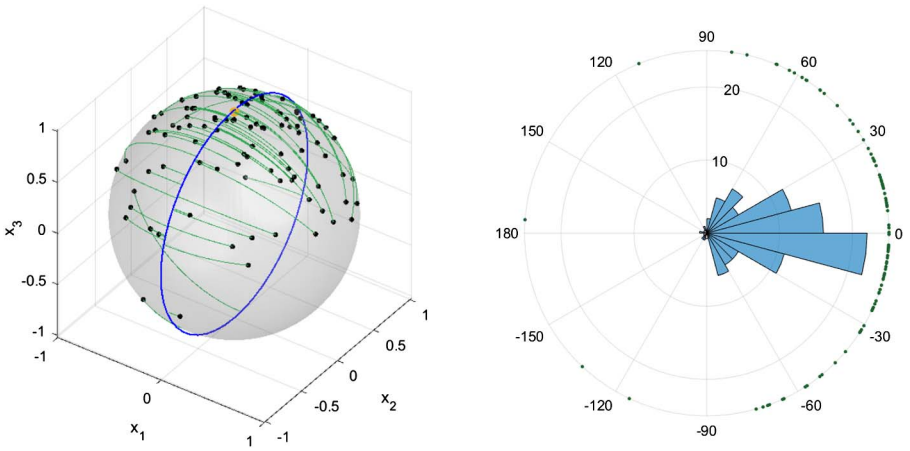

FIG 1. *(Left) A sample* $\mathbf{x}$ *of size* $n = 100$ *from* $vMF_3(\mathbf{e}_3, 3)$*, and its geodesic-projection* $P_\gamma(\mathbf{x})$ *onto a canonical geodesic* $\gamma(\mathbf{q}, \mathbf{v})$ *(blue great circle), where* $\mathbf{q} = \mathbf{e}_3$*,* $\mathbf{v} = \mathbf{e}_2$*. Also shown are the trajectories of the projection of each observation. (Right) A histrogram and dot plot for the projected score* $S_\gamma(\mathbf{x})$ *(shown in degrees).*

distribution in the areas of directional data, a natural question is whether the projected score of a vMF-distributed $\mathbf{x}$ is also a vM. We show that this is not the case.

The distribution of $S_\gamma(\mathbf{x})$ can be parameterized by p, κ and the deviation δ of γ from the mode $\boldsymbol{\mu}$, due to the rotational symmetry of the vMF distribution. The deviation is defined as

$$\delta = \rho(\boldsymbol{\mu}, \gamma) := \min_{\mathbf{y} \in \gamma(\mathbb{R})} \rho(\boldsymbol{\mu}, \mathbf{y}) \in [0, \pi/2]. \tag{3.1}$$

With respect to the mode $\boldsymbol{\mu}$ of the distribution, a geodesic γ is parameterized by $(\mathbf{q}, \mathbf{v})$, where $\mathbf{q}$ is the point on $\gamma(\mathbb{R})$ closest to $\boldsymbol{\mu}$, i.e., $\delta = \rho(\boldsymbol{\mu}, \mathbf{q})$, and $\mathbf{v} \in \gamma(\mathbb{R})$ is any of the two unit vectors orthogonal to $\mathbf{q}$. For $\delta < \pi/2$, such $\mathbf{q}$ is unique. The distribution of $S_\gamma(\mathbf{x})$ is invariant to the choice of $\mathbf{v}$, and to the choice of $\mathbf{q}$ if $\delta = \pi/2$. In any case, $\mathbf{v}'\boldsymbol{\mu} = 0$.

Among the geodesics, we pay special attention to the geodesics passing through the mean $\boldsymbol{\mu}$, i.e., the case $\delta = 0$. Any geodesic passing though $\boldsymbol{\mu}$ can be parameterized by $\boldsymbol{\mu}$ and $\mathbf{v} \in \mathbb{S}^{p-1}$ satisfying $\mathbf{v}'\boldsymbol{\mu} = 0$. These geodesics $\gamma(\boldsymbol{\mu}, \mathbf{v})$ will be referred to as *canonical geodesics*, hereafter. The distribution of $S_\gamma(\mathbf{x})$ is identical for any canonical geodesic γ, due to the rotational symmetry of the vMF distribution.

The opposite extreme is the case $\delta = \pi/2$, where the geodesic γ is as far as possible from the center of the distribution.

In the following, we inspect the distribution of $S_\gamma(\mathbf{x})$, $\mathbf{x} \sim \text{vMF}_p(\boldsymbol{\mu}, \kappa)$. The density function of $S_\gamma(\mathbf{x})$ takes a special form, and does not coincide with any known distributions. We denote $T \sim \text{PvM}(p, \kappa, \delta)$ for $T = S_\gamma(\mathbf{x})$, where δ is the

deviation of γ from $\boldsymbol{\mu}$, defined in (3.1). For the canonical geodesics with $\delta = 0$, we simply write $\mathrm{PvM}(p, \kappa)$ for $\mathrm{PvM}(p, \kappa, 0)$.

3.1. The density of $\mathbf{PvM}(p, \kappa, \delta)$

The density of $T = S_\gamma(\mathbf{x})$, with $\gamma = \gamma(\mathbf{q}, \mathbf{v})$, can be obtained by evaluating the joint density of $(X, Y) = (\mathbf{x}'\mathbf{q}, \mathbf{x}'\mathbf{v})$, then switching to the polar coordinates (R, T), where $R\cos T = X$ and $R\sin T = Y$. In general, there is no closed-form expression for the density of T, but we show that $T \sim \mathrm{PvM}(p, \kappa, \delta)$ is an infinite scale mixture of vM distributions.

The vM distribution $\mathrm{vM}(\mu, \kappa)$ is the special case of the vMF distribution for $p = 2$. Since $\mathbb{S}^1$ is the unit circle, the vM density is oftentimes given for angles rather than unit vectors as in (1.1). We say the angle $\theta \sim \mathrm{vM}(\mu, \kappa)$, for $\mu \in (-\pi, \pi]$ and $\kappa \geq 0$, if its density function is

$$f_{\mathrm{vM}}(\theta; \mu, \kappa) = c_2(\kappa) e^{\kappa \cos(\theta - \mu)}, \quad c_2(\kappa) = \frac{1}{2\pi I_0(\kappa)},$$

for $\theta \in (-\pi, \pi]$. We use the notation $\mathrm{vM}(\kappa) := \mathrm{vM}(0, \kappa)$ and $f_{\mathrm{vM}}(\theta; \kappa) := f_{\mathrm{vM}}(\theta; 0, \kappa)$ as we are only interested in mean-zero vM distributions.

Theorem 3.1. *Let $T \sim \mathrm{PvM}(p, \kappa, \delta)$ for $p \geq 3$, $\kappa \geq 0$ and $0 \leq \delta \leq \pi/2$. Then the density function of T is*

$$f_{\mathrm{PvM}}(t; p, \kappa, \delta) = \int_0^1 f_R(r) f_{\mathrm{vM}}(t; \kappa \cos(\delta) r) dr, \tag{3.2}$$

for $t \in (-\pi, \pi]$, where the mixing density f_R depends on p, κ, δ and is

$$f_R(r) = \frac{2}{I_\nu^*(\kappa)} I_0\{\kappa \cos(\delta) r\} r (1 - r^2)^{\nu - 1} I_{\nu-1}^*\{\kappa \sin(\delta) \sqrt{1 - r^2}\},$$

for $r \in (0, 1)$, where $\nu = (p - 2)/2$ and $I_\nu^*(z) = (\frac{z}{2})^{-\nu} I_\nu(z)$ for $z > 0$ and $I_\nu^*(0) = 1/\Gamma(\nu + 1)$.

For any $p \geq 3$, $\kappa > 0$ and $\delta \in [0, \pi/2)$, $f_{\mathrm{PvM}}(t)$ is symmetric about $t = 0$, strictly decreasing on $t > 0$, and has its unique mode at $t = 0$. These properties are inherited from the mean-zero vM distributions. In Fig. 2, an example of the joint density of (R, T) is displayed. If $\kappa = 0$, both $\mathrm{vMF}_p(\boldsymbol{\mu}, 0)$ and $\mathrm{PvM}(p, 0, \delta)$ are the uniform distributions on $\mathbb{S}^{p-1}$ and $\mathbb{S}^1$, respectively.

For the special case of $\delta = 0$ (that is, when γ is a canonical geodesic), we have $f_{\mathrm{PvM}}(t; p, \kappa) = f_{\mathrm{PvM}}(t; p, \kappa, 0) = \int_0^1 f_R(r) f_{\mathrm{vM}}(t; \kappa r) dr$, with a simpler expression of the mixing density

$$f_R(r) = \frac{2(\frac{\kappa}{2})^\nu}{I_\nu(\kappa)\Gamma(\nu)} I_0(\kappa r) r (1 - r^2)^{\nu - 1}.$$

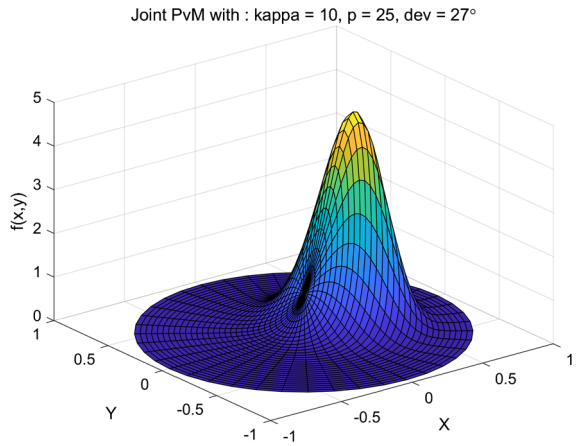

FIG 2. *The joint density of* $(X, Y) = (\mathbf{x}'\mathbf{q}, \mathbf{x}'\mathbf{v})$ *when* $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$. *Here,* $\delta = 27°, p = 25, \kappa = 10$. *In the polar coordinates* (R, T), *where* $R\cos T = X$ *and* $R\sin T = Y$, *each conditional distribution of* T *given* R *is a vM.*

Even in this special case, the integral involved has no closed form solution. Fortunately, it can be represented by the values of special functions I_ν and M_ν. Our next result is on the exact density function of $S_\gamma(\mathbf{x})$, when γ is a canonical geodesic.

Theorem 3.2. *Let* $\mathbf{x} \in \mathbb{S}^{p-1}$, $p \geq 3$, *follow* $\mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$ *for* $\kappa \geq 0$. *Suppose that the geodesic* $\gamma = \gamma(\mathbf{q}, \mathbf{v})$ *passes through* $\boldsymbol{\mu}$ *with* $\mathbf{q} = \boldsymbol{\mu}$. *Then the density function of the projection score* $S_\gamma(\mathbf{x})$ *is* $f_{\mathrm{PvM}}(t; p, \kappa) = C(p, \kappa)(1 + h(t))$ *for* $t \in (-\pi, \pi]$, *where*

$$h(t) = \begin{cases} \frac{\sqrt{\pi}\Gamma(\frac{p}{2})}{\left(\frac{|\kappa\cos t|}{2}\right)^{\frac{p-3}{2}}} \left(M_{\frac{p-1}{2}}(|\kappa\cos t|) + 2I_{\frac{p-1}{2}}(|\kappa\cos t|) \right), & \text{if } t \in (-\frac{\pi}{2}, \frac{\pi}{2}); \\ 0, & \text{if } t = \pm\frac{\pi}{2}; \\ \frac{\sqrt{\pi}\Gamma(\frac{p}{2})}{\left(\frac{|\kappa\cos t|}{2}\right)^{\frac{p-3}{2}}} M_{\frac{p-1}{2}}(|\kappa\cos t|), & \text{if } |t| \in (\frac{\pi}{2}, \pi], \end{cases} \tag{3.3}$$

and

$$C(p, \kappa) = \frac{\left(\frac{\kappa}{2}\right)^{\frac{p}{2}-1}}{2\pi(\frac{p}{2}-1)\Gamma(\frac{p}{2}-1)I_{\frac{p}{2}-1}(\kappa)}.$$

In Fig. 3, the density function f_{PvM} is plotted for a few choices of the dimension-concentration pair (p, κ). There, the closest vM density is overlaid as well. The precise meaning of "closest" is defined in Section 3.2. To human eye, the family of the density f_{PvM} is similar to the vM family.

For the opposite extreme case of $\delta = \pi/2$ where γ is as far from $\boldsymbol{\mu}$ as possible, all vM distributions in the mixture of (3.2) becomes the uniform distribution on $(-\pi, \pi]$, which in turn leads that $f_{\mathrm{PvM}}(t; p, \kappa, \pi/2)$ is uniform. A slightly more general statement is shown in Lemma 3.3.

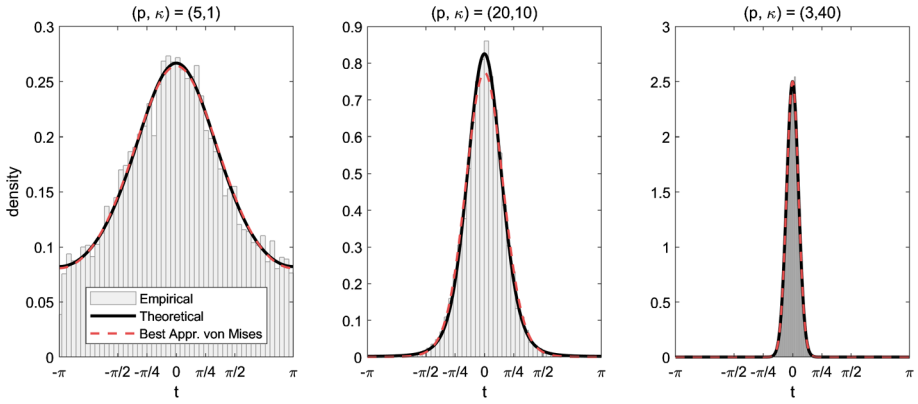

FIG 3. *Examples of the probability density function* $f_{\mathrm{PvM}}(\cdot; p, \kappa, 0)$ *of the projection score* $S_\gamma(\mathbf{x})$*, overlaid with the histogram of* $S_\gamma(\mathbf{x})$*, obtained from* $n = 10{,}000$ *random sample of* $\mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$*. The best approximating vM density, obtained by (3.4), shows that* f_{PvM} *is close to a vM density, but is not exactly the same with any vM density.*

Lemma 3.3. *Let* $\mathbf{x}$ *be a random variable on* $\mathbb{S}^{p-1}$ *whose distribution is rotationally symmetric about* $\boldsymbol{\mu}$*. Suppose that* $\mathbf{q}, \mathbf{v}$ *of* $\gamma(\mathbf{q}, \mathbf{v})$ *satisfy* $\mathbf{q}'\boldsymbol{\mu} = \mathbf{v}'\boldsymbol{\mu} = 0$*. Then* $S_\gamma(\mathbf{x}) \sim$ *Uniform*$(-\pi, \pi]$*.*

As κ increases, f_{PvM} becomes more concentrated. On the other hand, the higher the dimension p, the more dispersed f_{PvM}. As the deviation δ increases, the distribution of $S_\gamma(\mathbf{x})$ becomes more dispersed toward the uniform. This is graphically shown in Fig 4, in which the density of $S_\gamma(\mathbf{x})$ is obtained by the numerical integration of $f(r, t) = f_R(r) f_{\mathrm{vM}}(t; \kappa \cos(\delta) r)$ over $r \in (0, 1)$.

3.2. Comparisons to the Jones–Pewsey distribution family

Although f_{PvM} looks close to a vM density, it is not equal to any vM density. We show this for a more general distribution family, which includes the vM as a special case.

The Jones–Pewsey distribution (Jones and Pewsey, 2005) is a general three-parameter family of symmetric circular distributions. The density function of the Jones–Pewsey distribution, with location parameter μ, concentration parameter $\tilde{\kappa} \geq 0$, and a shape parameter $\phi \in \mathbb{R}$, is given by $g_{\mathrm{JP}}(\theta) = \omega(\tilde{\kappa}, \phi)\{1 + \tanh(\tilde{\kappa}\phi)\cos(\theta - \mu)\}^{1/\phi}$, for $\theta \in (-\pi, \pi)$, where $\omega(\tilde{\kappa}, \phi)$ is the normalizing constant. The Jones–Pewsey distribution includes the vM distribution, the Cardioid distribution, the wrapped Cauchy distribution and the circular t-distribution (Shimizu and Iida, 2002) as special cases; see Ley and Verdebout (2017) for a discussion on the Jones–Pewsey distribution.

The following result shows, in particular, that there is no vM distribution whose density function equals $f_{\mathrm{PvM}}(t; p, \kappa)$.

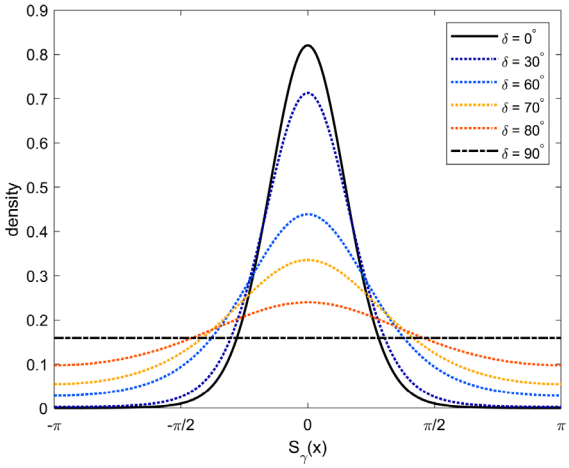

FIG 4. *Density functions of* $\mathrm{PvM}(p,\kappa,\delta)$ *for* $(p,\kappa)=(3,5)$ *and varying* $\delta=0$ *(solid), 30, 60, 70, 80 (dotted) and 90 (dashed) degrees.*

Theorem 3.4. *For any given* $p\geq 3$ *and* $\kappa\in(0,\infty)$*, there is no Jones–Pewsey distribution whose density function equals* $f_{\mathrm{PvM}}(t;p,\kappa)$*.*

Note that if $\kappa=0$ or ∞, then $f_{\mathrm{PvM}}(t)$ is the uniform distribution or the point mass at 0, respectively.

The discrepancy of the $\mathrm{PvM}(p,\kappa,\delta)$, $\delta\geq 0$, with the vM distribution is confirmed numerically. In Fig. 5, we compare f_{PvM} with the closest vM density function for a fine grid of (p,κ) and for $\delta=0,30^\circ,60^\circ$. The closest vM density is defined by $\mathrm{vM}(\tilde{\kappa}_0)$ for

$$\tilde{\kappa}_0=\arg\min_{\tilde{\kappa}} H(f_{\mathrm{PvM}}(\cdot;p,\kappa,\delta),f_{\mathrm{vM}}(\cdot;\tilde{\kappa})),\tag{3.4}$$

where $H(f,g)=\frac{1}{\sqrt{2}}\int(\sqrt{f(t)}-\sqrt{g(t)})^2dt$ is the Hellinger distance between two density functions f,g. The maximum Hellinger distance shown in the figure is about 0.0196.

While $\mathrm{PvM}(p,\kappa,\delta)$ is not exactly a vM distribution, they are virtually the same for the majority of possible (p,κ). To support this claim, we test the null hypothesis of $S_\gamma(\mathbf{x})\sim\mathrm{vM}(0,\tilde{\kappa})$ using a random sample of size n from $\mathrm{PvM}(p,\kappa,\delta)$. We use Kuiper (1960)'s goodness-of-fit test (adapted for the vM as in Section 6.2 of Pewsey, Neuhäuser and Ruxton (2013)), and $\tilde{\kappa}$ is given by the maximizer of the vM likelihood. We choose $(p,\kappa,\delta)=(20,10,0)$ as used in the middle panel of Fig. 3. For the sample size up to $n=1000$, the power of the test is less than 6% (numerically evaluated with significance level 5%), meaning that in the usual situations where the sample size is at most hundreds one cannot distinguish $\mathrm{PvM}(p,\kappa)$ with a vM distribution.

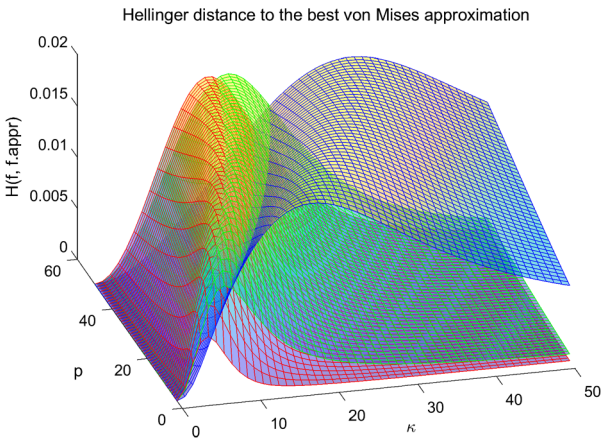

FIG 5. *The Hellinger distance between f_{PvM} and the closest vM density for various values of p, κ and for $\delta = 0$ (red), 30° (green) and 60° (blue).*

4. Approximations to vM distributions

In this section, we provide various approximations of $\mathrm{PvM}(p, \kappa, \delta)$ by the vM distributions, to support our claim that PvM is nearly vM. Asymptotic results in this section will also be useful in our discussion of projection pursuit, in Section 5. Since a $\mathrm{PvM}(p, \kappa, \delta)$ has zero mean, the approximating $\mathrm{vM}(\mu, \tilde{\kappa})$ has $\mu = 0$. We identify the concentration parameter $\tilde{\kappa}$ as a function of (p, κ, δ).

4.1. The case of canonical geodesics

When γ is a canonical geodesic, i.e., $\delta = 0$, with respect to $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$, the exact density function of $S_\gamma(\mathbf{x}) \sim \mathrm{PvM}(p, \kappa)$ is derived in Theorem 3.2. We approximate the density function $f_{\mathrm{PvM}}(t; p, \kappa)$ by a vM density under various asymptotic scenarios regarding the parameters of the vMF distribution. Specifically, we consider high concentration ($\kappa \to \infty$), low concentration ($\kappa \to 0$), high dimension ($p \to \infty$), and a combination of high-dimension and high-concentration where both $p, \kappa \to \infty$ but $p/\kappa \in (0, \infty)$ is fixed.

The big Theta Θ notation is used to state our finding. Under the asymptotic regime $\nu \to \infty$, we say $f_\nu = \Theta(g_\nu)$ if $\lim_{\nu\to\infty} f_\nu / g_\nu = c \neq 0$. For example, for $\alpha \in \mathbb{R}$ fixed, $f_\nu = \Theta(\nu^{-\alpha})$ if and only if $\lim_{\nu\to\infty} \nu^\alpha f_\nu < \infty$ and for any $\epsilon > 0$, $\lim_{\nu\to\infty} \nu^{\alpha-\epsilon} f_\nu = 0$. The asymptotic regime may be sometimes $\nu \to 0$, and the definition of Θ is changed to reflect the appropriate limit.

We summarize our finding in Theorem 4.1. There, $\phi(\cdot)$ is the density function of the standard normal distribution. We say θ follows the Cardioid distribution (centered at $\mu = 0$) with parameter $\rho \in (0, \frac{1}{2})$ if its density is $f_{\mathrm{Cardioid}}(t; \rho) = \frac{1}{2\pi}(1 + 2\rho \cos t)$ on $t \in (-\pi, \pi)$. It is well-known that the vM distribution is approximated by the normal distribution, for high concentration, and by the

Cardioid distribution for low concentration; see Kent (1978) and Mardia and Jupp (2000).

Theorem 4.1. *For $T \sim \mathrm{PvM}(p,\kappa)$, let $f_\kappa(s;p,\kappa)$ be the density function of $\sqrt{\kappa}T$. Similarly, for $\theta \sim \mathrm{vM}(0,\tilde{\kappa})$, let $g_\kappa(s;\tilde{\kappa})$ be the density function of $\sqrt{\kappa}\theta$. Here, κ and $\tilde{\kappa}$ need not be the same.*

(i) (High concentration) For any fixed $p \geq 3$ and for any $s \in \mathbb{R}$, as $\kappa \to \infty$,

$$f_\kappa(s;p,\kappa) = g_\kappa(s;\kappa) + \Theta(\kappa^{-1}) = \phi(s) + \Theta(\kappa^{-1}).$$

Informally, we write $\mathrm{PvM}(p,\kappa) \approx \mathrm{vM}(0,\kappa) \approx N(0,\kappa^{-1})$ as $\kappa \to \infty$.

(ii) (Low concentration) Let $\tilde{\kappa}_l = B(\frac{p}{2},\frac{1}{2})\kappa/2$, where $B(\cdot,\cdot)$ is the beta function. For any fixed $p \geq 3$ and for any $t \in (-\pi,\pi)$, as $\kappa \to 0$,

$$f_{\mathrm{PvM}}(t;p,\kappa) = f_{\mathrm{vM}}(t;\tilde{\kappa}_l) + \Theta(\kappa^2) = f_{\mathrm{Cardioid}}(t;\tilde{\kappa}_l/2) + \Theta(\kappa^2).$$

(iii) (high dimension) Let $\tilde{\kappa}_d = \kappa\sqrt{\frac{\pi}{2(p-1)}}$. For any fixed $\kappa > 0$ and for any $t \in (-\pi,\pi)$, as $p \to \infty$,

$$f_{\mathrm{PvM}}(t;p,\kappa) = f_{\mathrm{vM}}(t;\tilde{\kappa}_d) + \Theta(p^{-1}) = f_{\mathrm{Cardioid}}(t;\tilde{\kappa}_d/2) + \Theta(p^{-1}).$$

(iv) (High concentration, high dimension) Let $\tilde{\kappa}_b = \kappa\frac{\lambda}{1+\sqrt{1+\lambda^2}}$, where $\lambda = \kappa/u$ and $u = (p-1)/2$, and $\varsigma_\lambda^2 = \frac{\kappa}{\tilde{\kappa}_b} = \frac{1+\sqrt{1+\lambda^2}}{\lambda}$. Suppose $\kappa \to \infty$ and $p \to \infty$ simultaneously while $\lambda = \kappa/u \in (0,\infty)$ is fixed. Then, for any $s \in \mathbb{R}$,

$$f_\kappa(s;p,\kappa) = g_\kappa(s;\tilde{\kappa}_b) + \Theta(p^{-1}) = \frac{1}{\varsigma_\lambda}\phi\left(\frac{s}{\varsigma_\lambda}\right) + \Theta(p^{-1}).$$

Informally, we write $\mathrm{PvM}(p,\kappa) \approx \mathrm{vM}(0,\tilde{\kappa}_b) \approx N(0,\tilde{\kappa}_b^{-1})$ as $\kappa, p \to \infty$ and λ fixed.

An immediate implication of Theorem 4.1 is that for $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu},\kappa)$, and for a canonical geodesic γ, $S_\gamma(\mathbf{x}) \Rightarrow \mathrm{Unif}(-\pi,\pi)$ as $\kappa \to 0$ or as $p \to \infty$. Here and throughout, the notation $\Rightarrow$ stands for the convergence in distribution. This result is consistent with the high-dimensional asymptotic results in the literature. In particular, it is shown in Watson (1988) and Dryden (2005) that for a fixed κ, $\mathrm{vMF}_p(\boldsymbol{\mu},\kappa)$ converges to the uniform distribution on $\mathbb{S}^\infty$ as $p \to \infty$. The uniform distribution on $\mathbb{S}^{p-1}$ is a special case of the vMF distribution with $\kappa = 0$, and $\mathrm{Unif}(-\pi,\pi)$ is a special case of the PvM distribution with $\kappa = 0$. Thus, the approximation by $\mathrm{vM}(\tilde{\kappa}_l)$ in the low concentration setting and the high-dimensional approximation $\mathrm{vM}(\tilde{\kappa}_d)$ are bound to be similar in the high dimensional situations. In fact, $\tilde{\kappa}_d \approx \tilde{\kappa}_l$ for large p; for $u = (p-1)/2$, Gautschi's inequality (Qi, 2010) gives

$$1 < \frac{\tilde{\kappa}_d}{\tilde{\kappa}_l} = u^{-\frac{1}{2}}\frac{\Gamma(u+\frac{1}{2})}{\Gamma(u+1)} < \sqrt{1+\frac{1}{2u}}.$$

In the opposite case where $\kappa \to \infty$, the underlying $\mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$ converges to the point mass at $\boldsymbol{\mu}$. The rate of convergence is exactly $\kappa^{-1/2}$ and $\mathrm{PvM}(p, \kappa)$ is approximately vM or normal. In the high-dimensional case, if the concentration parameter κ increases at the same rate as p, then Case (iv) in Theorem 4.1 leads that for $\lambda = 2\kappa/p \in (0, \infty)$,

$$\left(\frac{\sqrt{\lambda^2 + 1} - 1}{\lambda} \kappa \right)^{\frac{1}{2}} S_\gamma(\mathbf{x}) \Rightarrow N(0, 1),$$

as $\kappa \to \infty$. We remark that if κ increases at the rate $\kappa = \sqrt{p}\lambda$ for $\lambda \in (0, \infty)$, the distribution of $S_\gamma(\mathbf{x})$ neither diverges to the uniform nor degenerates to a point mass. This special case will be discussed in the general setting where $\delta \geq 0$ in Section 4.2.

4.2. The case of general geodesics

Approximation of $\mathrm{PvM}(p, \kappa, \delta)$ in the general case where $\delta \in [0, \pi/2]$ is discussed in this section. Since the exact density function is not available for $\delta > 0$, the results derived here are weaker than the corresponding results for $\delta = 0$ in Section 4.1, in the sense that the approximation error is not evaluated in most cases. An exception is the low-concentration case $\kappa \to 0$ (while other parameters are fixed), and an extension of Theorem 4.1 (ii) is given. We also inspect two different asymptotic regimes of high-dimensional cases $p \to \infty$ and $p, \kappa \to \infty$ at the same rate. It is seen that $\mathrm{PvM}(p, \kappa, \delta)$ becomes either as dispersed as possible (when $p \to \infty$ and κ is fixed) or converges to 0 with the rate $\kappa^{-1/2}$ (when $p \to \infty$ and $\kappa = \Theta(p)$). An interesting result is obtained when we consider the case $\kappa = \Theta(p^{1/2})$. As $p \to \infty$, $\mathrm{PvM}(p, \kappa, \delta)$ does not degenerate in the limit, but converges to the projected normal distribution.

In the following, we discuss the four asymptotic scenarios of approximating $\mathrm{PvM}(p, \kappa, \delta)$.

4.2.1. Low concentration approximation

The distribution of $S_\gamma(\mathbf{x})$, $\mathrm{PvM}(p, \kappa, \delta)$, is a mixture of $\mathrm{vM}(\kappa \cos(\delta) R)$ over $R \in (0, 1)$, as given in Theorem 3.1. In the low-concentration asymptotic scenario of $\kappa \to 0$, all concentration parameters of the vM converge to zero, but the mixing density f_R does not degenerate. Nevertheless, a mixture of vM(0)'s is simply vM(0), which is the uniform distribution. A much stronger statement is given in Theorem 4.2, where we provide an approximation of $\mathrm{PvM}(p, \kappa, \delta)$ by $\mathrm{vM}(\tilde{\kappa})$, and its approximation error.

Theorem 4.2. *Let* $\tilde{\kappa} = \tilde{\kappa}(p, \kappa, \delta) = B(\frac{p}{2}, \frac{1}{2})\kappa \cos \delta/2$. *For any* $p \geq 3$, $\delta \in [0, \pi/2)$, *and for any* $t \in (-\pi, \pi]$, *as* $\kappa \to 0$,

$$f_{\mathrm{PvM}}(t; p, \kappa, \delta) = f_{\mathrm{vM}}(t; \tilde{\kappa}) + \Theta(\kappa^2).$$

Note that the above results coincides with Theorem 4.1 (ii) for $\delta = 0$. It is conjectured that, for $\delta > 0$, the approximations in the other asymptotic regimes are simply those in Theorem 4.1 with κ replaced by $\kappa \cos\delta$. While the conjecture seems true for the case $p, \kappa \to \infty$ discussed in Theorem 4.3 later, a rigorous proof has not been obtained.

In our proof of Theorem 4.2, detailed in the Appendix, we provide a uniform bound of the density function f_{PvM}:

$$f^{(-)}(t) \le f_{\mathrm{PvM}}(t) \le f^{(+)}$$

over $t \in (-\pi, \pi]$. The functions $f^{(-)}(t)$ and $f^{(+)}$ are easier to handle as the integrals involved have closed-form expressions, and both converge to the uniform density as $\kappa \to 0$.

4.2.2. High-dimensional approximations

We consider three asymptotic regimes: κ is fixed, $\kappa = \Theta(p)$ and $\kappa = \Theta(\sqrt{p})$ as $p \to \infty$.

High-dimension with κ fixed Using the joint density function of $(R, S_\gamma(\mathbf{x}))$ in Theorem 3.1 and the large-order asymptotic expression of I_ν in (A.9), it can be shown that $R \to 0$ in probability, and furthermore,

$$(\sqrt{p}R, S_\gamma(\mathbf{x})) \Rightarrow (X, U), \quad \text{as } p \to \infty, \tag{4.1}$$

where $X \sim \mathrm{Rayleigh}(1)$ and $U \sim \mathrm{Uniform}(-\pi, \pi)$ are independent. A proof of (4.1) is given in Appendix Section B.3. This result is not entirely new and can be derived from Watson (1988).

Watson investigated the vMF distribution in high-dimensional spheres, and in particular the orthogonal projections of a vMF-distributed random directions onto a subspace. For the projection onto the subspace spanned by $\gamma(\mathbf{q}, \mathbf{v})$, let $\mathbf{P} = [\mathbf{q}, \mathbf{v}]$ be the orthogonal frame of the subspace. Watson (1988) showed that

$$\sqrt{p}\mathbf{P}'\mathbf{x} \Rightarrow N_2(\mathbf{0}, I_2) \quad \text{as} \quad p \to \infty;$$

see also Dryden (2005). Since the polar angle $S_\gamma(\mathbf{x})$ is invariant to positive simultaneous scaling of $\mathbf{P}'\mathbf{x} = (\mathbf{q}'\mathbf{x}, \mathbf{v}'\mathbf{x})$, Lemma 2.1 and the continuous mapping theorem give

$$S_\gamma(\mathbf{x}) \Rightarrow \mathrm{Uniform}(-\pi, \pi) \quad \text{as} \quad p \to \infty.$$

Comparing to Theorem 4.1 (iii), where $\delta = 0$, we conjecture that $\mathrm{PvM}(p, \kappa, \delta) \approx \mathrm{vM}(\tilde{\kappa}_d)$ as $p \to \infty$ where $\tilde{\kappa}_d = \kappa \cos\delta \sqrt{\frac{\pi}{2(p-1)}}$.

High-dimension, high-concentration with $\kappa = \Theta(p)$ Assume that both p and κ increases at the same rate. If the deviation δ of the geodesic γ from the

mode $\boldsymbol{\mu}$ of $\mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$ is not its maximum, i.e. $\delta < \pi/2$, then $S_\gamma(\mathbf{x})$ degenerates as $p \to \infty$ and can be well approximated by either the vM or normal distribution. In the following we show the convergence of a scaled $S_\gamma(\mathbf{x})$ in the limit $p, \kappa \to \infty$, while $\kappa/p \to \lambda/2 \in (0, \infty)$.

In the general geodesic case where $\delta > 0$, the distribution of $S_\gamma(\mathbf{x})$ is only available as the mixture of vM distributions $\mathrm{vM}(\kappa \cos(\delta) R)$. It turns out as $p \to \infty$, the mixing random variable R concentrates at

$$r_0 = \frac{\lambda \cos\delta}{1 + \sqrt{1 + \lambda^2}},$$

shown in Lemma B.7 in Appendix.

Informally, $S_\gamma(\mathbf{x})$ is the mixture of $\mathrm{vM}(\kappa \cos(\delta) r_0)$, with weight approaching 1, and the rest with weight approaching 0, as $p \to \infty$. This entails that $S_\gamma(\mathbf{x}) \sim \mathrm{vM}(\kappa \cos(\delta) r_0)$ in the limit, but because κ increases as well, the vM is again approximated by the normal distribution. This is shown next.

Theorem 4.3. *Let $u = p/2 - 2$ for $p \geq 5$ and $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$. Suppose both u and κ increase while $\lambda = \kappa/u \in (0, \infty)$ is fixed, and $\delta = \rho(\boldsymbol{\mu}, \gamma)$ is fixed.*

(i) If $\delta \in [0, \pi/2)$, then for

$$\tilde{\kappa} = \kappa \cos^2(\delta) \frac{\lambda}{1 + \sqrt{1 + \lambda^2}},$$

$\sqrt{\tilde{\kappa}} S_\gamma(\mathbf{x}) \Rightarrow N(0, 1)$ as $p \to \infty$.

(ii) If $\delta = \pi/2$, then $S_\gamma(\mathbf{x}) \Rightarrow \mathrm{Uniform}(-\pi, \pi)$, as $p \to \infty$.

Remark 1. In Theorem 4.3, we used $u = (p-4)/2$ as it is the easiest form to handle, but the conclusion remains the same if $u = (p-1)/2$ or $p/2$ is used instead.

For the special case of a canonical geodesic γ with $\cos\delta = 1$, the statement of the above theorem can be obtained from Theorem 4.1 (iv). For this special case, Theorem 4.1 (iv) is in fact a stronger statement than Theorem 4.3, as the rate of convergence $\Theta(p^{-1})$ is not available in the latter.

Theorem 4.3, together with Lemma B.2, implies that if both p and κ are large, $\mathrm{PvM}(p, \kappa, \delta)$ is well approximated by $\mathrm{vM}(\tilde{\kappa})$ and by $N(0, \tilde{\kappa}^{-1})$.

High-dimension, moderately large $\kappa = \Theta(\sqrt{p})$ The distribution of $S_\gamma(\mathbf{x})$ does not degenerate when κ increases at the rate of $\sqrt{p}$. The case $p \to \infty$ while $\kappa = \sqrt{p}\lambda$ is investigated by Watson (1988), in which Watson showed that the orthogonal projection of $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$ onto the subspace spanned by $\mathbf{P} = [\mathbf{p}, \mathbf{v}]$ converges to a normal distribution when scaled by $\sqrt{p}$. We show in Theorem 4.4 that the polar angle $S_\gamma(\mathbf{x})$ of the projection score $p^{1/2}\mathbf{P}'\mathbf{x}$ converges to the distribution of polar angle of the limiting 2-variate normal distribution, called the projected normal distribution.

In the theorem, we in addition allow the geodesic $\gamma = \gamma(\mathbf{p}, \mathbf{v})$ to be random, but require that for a given $\delta \in [0, \pi/2]$, $\rho(\boldsymbol{\mu}, \gamma) \to \delta$ in probability as $p \to \infty$. This will be useful in Section 5.

The projected normal distribution, sometimes referred to as angular Gaussian or offset normal distribution, is obtained by projecting a Gaussian random vector on $\mathbb{R}^q$ on $\mathbb{S}^{q-1}$ (in our case, $q = 2$), and does not coincide with the vMF distribution. We write $\mathrm{PN}_2(\boldsymbol{\mu}, \mathbf{I}_2)$ for the distribution of the polar angle of $N_2(\boldsymbol{\mu}, \mathbf{I}_2)$.

Theorem 4.4. *Consider a sequence* $\{\mathbf{x}_p : p = 3, 4, \ldots\}$ *of random directions, where* $\mathbf{x}_p \sim \mathrm{vMF}_p(\boldsymbol{\mu}_p, \kappa_p)$, *where for a* $\lambda > 0$, $\kappa_p = \sqrt{p}\lambda$. *For each* p, *let* γ_p *be a random geodesic parameterized with respect to* $\boldsymbol{\mu}_p$, *that is,* $\gamma_p = \gamma(\mathbf{q}, \mathbf{v})$ *and* $\Delta_p := \cos^{-1}(\mathbf{q}'\boldsymbol{\mu}_p) = \rho(\boldsymbol{\mu}_p, \gamma)$ *and* $\mathbf{v}'\boldsymbol{\mu}_p = 0$. *Assume that* γ_p *and* $\mathbf{x}_p$ *are independent, and for a constant* $\delta \in [0, \pi/2]$, $\Delta_p \to \delta$ *in probability as* $p \to \infty$. *Then as* $p \to \infty$,

$$\sqrt{p}(\mathbf{q}'\mathbf{x}_p, \mathbf{v}'\mathbf{x}_p)' \Rightarrow N_2((\lambda \cos\delta, 0)', \mathbf{I}_2),$$

and

$$S_\gamma(\mathbf{x}_p) \Rightarrow \mathrm{PN}_2((\lambda \cos\delta, 0)', \mathbf{I}_2).$$

The density function of $\mathrm{PN}_2((\lambda\cos\delta, 0)', \mathbf{I}_2)$ is

$$f_{\mathrm{PN}}(\theta; \lambda_\delta) = \frac{1}{2\pi} e^{-\lambda_\delta^2/2} \left\{ 1 + \frac{\Phi(\lambda_\delta \cos\theta)}{\phi(\lambda_\delta \cos\theta)} \lambda_\delta \cos\theta \right\},$$

for $\theta \in [-\pi, \pi)$, where $\lambda_\delta = \lambda\cos\delta$, and Φ and ϕ are the distribution function and the density function of the standard normal distribution, respectively (Watson, 1983; Pukkila and Rao, 1988; Presnell, Morrison and Littell, 1998). The density f_{PN} is unimodal, symmetric about zero, and exhibits heavier tails than the vM. Albeit the difference, the density $f_{\mathrm{PN}}(\theta; \lambda_\delta)$ is well approximated by vM distributions when λ_δ is low or high. In particular, for high concentration case where $\lambda_\delta \to \infty$, we have

$$\mathrm{PN}_2((\lambda_\delta, 0)', I_2) \approx \mathrm{vM}(\lambda_\delta^2) \approx N(0, \lambda_\delta^{-2}); \tag{4.2}$$

for low concentration case where $\lambda_\delta \to 0$, we have for $\tilde{\kappa} = \sqrt{\pi/2}\lambda_\delta$,

$$\mathrm{PN}_2((\lambda_\delta, 0)', I_2) \approx \mathrm{vM}(\tilde{\kappa}) \approx \mathrm{Cardioid}(\tilde{\kappa}/2). \tag{4.3}$$

Thus, if $p \to \infty$, $\kappa = \sqrt{p}\lambda$ and $\lambda\cos\delta$ is large, then $\mathrm{PvM}(p, \kappa, \delta)$ can be approximated by $\mathrm{vM}(p^{-1}\kappa^2\cos^2\delta)$. On the other hand, for $p \to \infty$, $\kappa = \sqrt{p}\lambda$ but $\lambda\cos\delta \approx 0$, we have $\mathrm{PvM}(p, \kappa, \delta) \approx \mathrm{vM}(\sqrt{\pi/(2p)}\kappa\cos\delta)$. Note that the latter vM approximation of PN is asymptotically equivalent to the high-dimensional approximation by $\mathrm{vM}(\sqrt{\frac{\pi}{2(p-1)}}\kappa)$ given in Theorem 4.1 (iii), for $\delta = 0$.

The assertions (4.2) and (4.3) are formally stated in Lemma 4.5.

Lemma 4.5. *For* $T \sim \mathrm{PN}_2((\lambda, 0)', I_2)$, *let* $f_{\mathrm{PN}}(t; \lambda)$ *and* $f_\lambda(s)$ *be the density function of* T *and* λT, *respectively. Similarly, for* $\theta \sim \mathrm{vM}(0, \tilde{\kappa})$, *let* $g_\lambda(s; \tilde{\kappa})$ *be the density function of* $\lambda\theta$.

TABLE 1
Asymptotic approximations of $\mathrm{PvM}(p,\kappa,\delta)$. (∗*: Shown only for* $\delta = 0$.)

Scenario	$\tilde{\kappa}$ of $\mathrm{vM}(\tilde{\kappa})$	Relevant result
p fixed, $\kappa \to \infty^*$	$\kappa \cos\delta$	Theorem 4.1 (i)
p fixed, $\kappa \to 0$	$B(\frac{p}{2}, \frac{1}{2})\kappa \cos\delta/2$	Theorems 4.1 (ii) and 4.2
$p \to \infty$, κ fixed*	$\kappa \cos\delta \sqrt{\frac{\pi}{2(p-1)}}$	Theorem 4.1 (iii)
$p \to \infty$, $\kappa \to \infty$, $\kappa/p \to \lambda/2$	$\kappa \cos^2\delta \frac{\lambda}{1+\sqrt{1+\lambda^2}}$	Theorems 4.1 (iv) and 4.3
$p \to \infty$, $\kappa \to \infty$, $\kappa/\sqrt{p} \to \lambda$	$\mathrm{PN}_2((\lambda\cos\delta, 0)', I_2)$	Theorem 4.4

(i) (High concentration) For any $s \in \mathbb{R}$, *as* $\lambda \to \infty$,

$$f_\lambda(s) = g_\lambda(s; \lambda^2) + \Theta(\lambda^{-2}) = \phi(s) + \Theta(\lambda^{-2}).$$

(ii) (Low concentration) Let $\tilde{\lambda} = \sqrt{\pi/2}\lambda$. *Then for any* $\lambda > 0$ *and* $t \in (-\pi, \pi)$, *as* $\lambda \to 0$,

$$f_{\mathrm{PN}}(t; \lambda) = f_{\mathrm{vM}}(t; \tilde{\lambda}) + \Theta(\lambda^2) = f_{\mathrm{Cardioid}}(t; \tilde{\lambda}/2) + \Theta(\lambda^2).$$

4.3. Summary and comparison

In Sections 4.1 and 4.2, several approximations of $\mathrm{PvM}(p,\kappa,\delta)$ are identified and justified in various asymptotic scenarios. A summary is given in Table 1.

A natural question to ask is that for a given (p,κ), which approximation works the best? We attempt to answer this question numerically. For each $p = 3,\ldots,50$, a fine grid of $\kappa \in (0, 70]$, and $\delta \in \{0, 30°, 60°, 80°\}$, we compare $f_{\mathrm{PvM}}(\cdot; p, \kappa, \delta)$ with each of the five approximations given in Table 1, and record the closest approximation to $\mathrm{PvM}(p,\kappa,\delta)$ in terms of the Hellinger distance.

The result is shown in Fig. 6. There, the high p, large κ approximation (the 4th row in Table 1) is shown to be superior than the others for most cases. When δ increases, the PvM distribution becomes less concentrated, and the low-concentration (small κ) approximation becomes better. We note that the high-dimensional approximations are similar to the low-concentration approximations for most cases of (p,κ). The approximation by the projected normal is seen to be the closest roughly for the cases $\kappa < 2p$. However, the difference between the approximating projected normal density and the other approximating vM densities, especially the small κ and high p approximations, are not large (in fact thin). These claims are numerically confirmed in Appendix C, in which we provide a more careful numerical comparison on the quality of approximations.

5. An application to projection pursuit of directional data

Projection pursuit (Friedman and Tukey, 1974) is a technique for exploration of multivariate analysis, aiming at finding low-dimensional "interesting" projections of high-dimensional data in $\mathbb{R}^p$. A direction $\mathbf{u} \in \mathbb{R}^p$ is deemed interesting if the distribution of the orthogonal projection score $\mathbf{x}'\mathbf{u}$ is as far from Gaussian

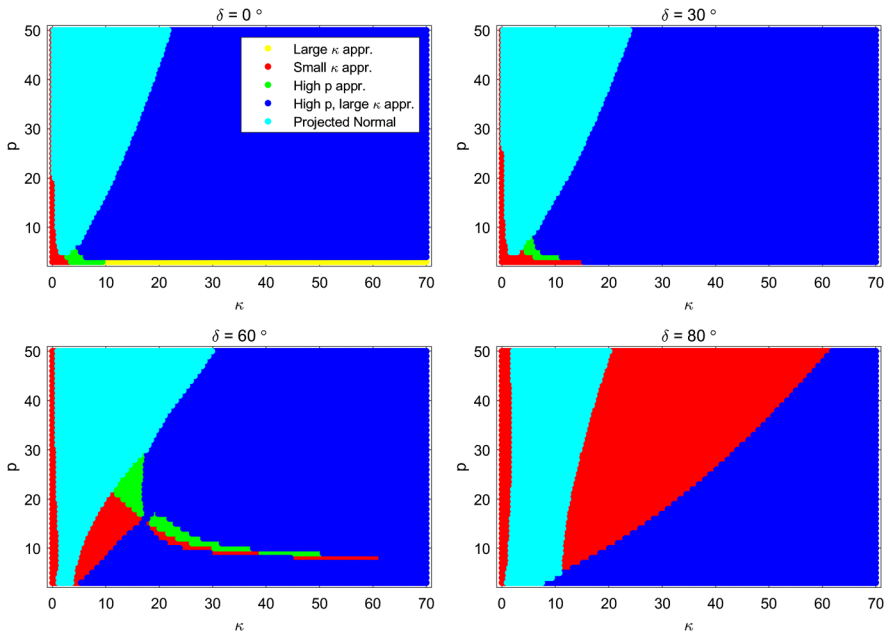

FIG 6. *Regions of best approximations in terms of the smallest Hellinger distance.*

as possible (Friedman, 1987; Bickel, Kur and Nadler, 2018). Such a non-normal distribution is an evidence of potential clusters or outliers in the multivariate data.

For directional data on $\mathbb{S}^{p-1}$, we propose to seek the geodesic γ such that the projection score $S_\gamma(\mathbf{x})$ is as non-vM as possible. As claimed in preceding sections, the projection score of a vMF-distributed random direction is approximately vM-distributed. Since any vMF distribution is rotationally symmetric about its mode, no geodesic is more interesting than any other geodesics. On the other hand, if the data exhibits an interesting mode of variation such as clusters, the projection score onto the geodesic passing through two clusters is bimodal, and strongly non-vM.

We demonstrate that a measure of non-vM-ness is useful in detecting clusters and outliers. In the following two examples, conformity to the vM distributions is measured by the p-value of the Kuiper's test with respect to the vM family, using a sample of projection scores $S_{\gamma_i}(\mathbf{x})$.

Example 1. Consider a location mixture of vMF, given by $\mathbf{x} \sim \frac{1}{2}\mathrm{vMF}_3(\boldsymbol{\mu}_1, \kappa) + \frac{1}{2}\mathrm{vMF}_3(\boldsymbol{\mu}_2, \kappa)$, where $\rho(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \pi/4$ and $\kappa = 50$. A sample of size $n = 100$ from this model is plotted in Fig. 7. While we have chosen to use $p = 3$ for presentational purposes, exploration of directional data for $p > 3$ by visualization is limited. We compare two geodesics γ_1 and γ_2, where γ_1 passes through both $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and γ_2 stems from the sample Fréchet mean of data, the minimizer $\boldsymbol{\mu}$ of $\sum_{i=1}^n \{\rho(\mathbf{x}_i, \boldsymbol{\mu})\}^2$, and is orthogonal to γ_1. If the data were in Euclidean

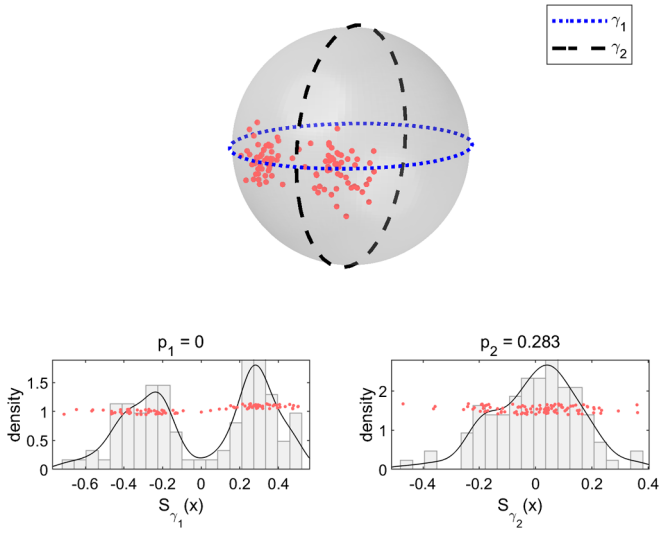

FIG 7. *A sample from a location mixture of vMF distributions, overlaid with geodesics* γ_1 *and* γ_2 *of interest. The strongly non-vM-distributed* $S_{\gamma_1}(\mathbf{x})$ *is related to the usefulness of* γ_1 *in exploratory data analysis.*

space, then γ_1 corresponds to the Fisher's discriminant direction. The clustered nature of the population appears along the geodesic γ_1 whose projection score is strongly non-vM (small p-value); see the bottom left panel of Fig. 7. On the other hand, the projection onto γ_2 is vM-distributed, and is deemed uninteresting.

Example 2. We consider a contaminated $\text{vMF}_3(\boldsymbol{\mu}, 50)$, where 10 percent of the observations are located 90 degrees away from $\boldsymbol{\mu}$. A sample of size $n = 50$ is plotted in Fig. 8. Two geodesics are compared as well: γ_2 passes through the two clusters and γ_1 stems from the sample Fréchet mean of the data and is orthogonal to γ_2. It can be checked that the outliers appear on $S_{\gamma_2}(\mathbf{x})$ (which is strongly non-vM), while the uninteresting γ_1 has its scores distributed not significantly different from a vM distribution.

While the above two examples concern a low-dimensional situation, the departure from the vM distribution is a viable measure of interestingness in high-dimensional situations. We show that under the location mixture model, similar to the examples above, almost all geodesics γ are uninteresting, i.e., the projections are approximately vM. Therefore, a geodesic γ_0 with a non-vM projection score is certainly interesting.

In particular, we use the high-dimensional setting of Theorem 4.4, and choose the geodesic γ from the uniform distribution on the set of all geodesics on $\mathbb{S}^{p-1}$. Note that the uniformly distributed γ is given by choosing a random orthogonal 2-frame $(\mathbf{q}, \mathbf{v})$. We also consider the case where the geodesic is confined to pass through the mean of the directional data, a common practice in dimension reduction. The result is easily derived from Theorem 4.4.

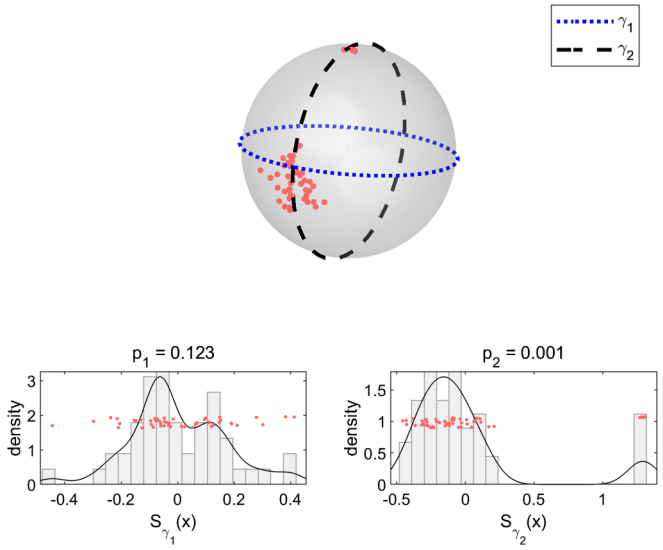

FIG 8. *A sample from a contaminated vMF distribution with outliers.*

Corollary 5.1. *Suppose that* $\mathbf{x}_p \sim \pi_1 \mathrm{vMF}_p(\boldsymbol{\mu}_1, \kappa_p) + (1-\pi_1)\mathrm{vMF}_p(\boldsymbol{\mu}_2, \kappa_p)$, $\kappa_p = \sqrt{p}\lambda$ *for some* $\lambda > 0$ *and* $\pi_1 \in (0,1)$.

(i) Suppose that $\gamma = \gamma(\mathbf{q}, \mathbf{v})$ *is chosen from the uniform distribution on the Stiefel manifold* $V_{2,p}$ *(consisting of 2-frames in* $\mathbb{R}^p$*). Then,*

$$S_\gamma(\mathbf{x}_p) \Rightarrow \mathrm{Uniform}(-\pi, \pi)$$

as $p \to \infty$.

(ii) Let $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$ *be any minimizer of* $\sum_{i=1}^2 \rho^2(\boldsymbol{\mu}_i, \boldsymbol{\mu})$. *Suppose that* $\rho(\boldsymbol{\mu}, \boldsymbol{\mu}_1) = \rho(\boldsymbol{\mu}, \boldsymbol{\mu}_2) \to d \in [0, \pi/2]$ *as* $p \to \infty$, *and that* $\mathbf{v}$ *is sampled from the uniform distribution on the* $(p-2)$ *dimensional unit sphere in the null space of* $\boldsymbol{\mu}$. *Then, for* $\gamma = \gamma(\boldsymbol{\mu}, \mathbf{v})$,

$$S_\gamma(\mathbf{x}_p) \Rightarrow \mathrm{PN}_2((\lambda \cos d, 0)', \mathbf{I}_2)$$

as $p \to \infty$.

The result above shows that as $p \to \infty$ for almost all choices of γ, the corresponding projection score is approximately vM-distributed (either uniform or the projected normal). However, as demonstrated in Example 1, there certainly is an interesting direction: For the canonical geodesic γ with respect to both $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, i.e., $\gamma = \Gamma(\boldsymbol{\mu}_1 \to \boldsymbol{\mu}_2)$, $S_\gamma(\mathbf{x})$ converges to a location mixture of two projected normal distributions.

Corollary 5.1 also hints a use of "non-Gaussianity" as a projection index in high dimensions. Unlike the usual vector-valued data where all uninteresting projections are Gaussian, uninteresting geodesic projections of directional data can be uniformly distributed. The uniform distribution is a special case of vM,

but is strongly non-Gaussian. Thus, for directional data, non-normality of scores does not always indicate interesting patterns of data.

Remark 2. Extensions of Corollary 5.1 may be necessary in developing a projection pursuit for high-dimensional directional data. For example, instead of the 2-mixture of equal concentrations assumed in Corollary 5.1, consider a K-mixture of $\mathrm{vMF}_p(\boldsymbol{\mu}_k, \lambda_k\sqrt{p})$ for $K > 2$ with possibly different concentration parameters λ_k. If the number K of mixtures is fixed, then it can be checked that the assertion (i) of Corollary 5.1 still holds even if λ_k's are distinct. On the other hand, generalizing the assertion (ii) is tricky. With more than two components, the limiting distances from the component centers to the grand mean, $d_k := \lim_{p\to\infty} \rho(\boldsymbol{\mu}, \boldsymbol{\mu}_k)$, should be treated arbitrary (as opposed to $d_1 = d_2 = d$ in the $K = 2$ case). With potentially different concentration parameters λ_k, we conjecture that $S_\gamma(\mathbf{x}_p)$ converges to a scale mixture of $\mathrm{PN}_2((\lambda_k \cos d_k, 0)', \mathbf{I}_2)$. Such a situation in a low-dimensional setting was explored in Example 2. There, the location mixture $S_{\gamma_2}(\mathbf{x})$ is shown to be more severely non-vM compared to the scale mixture $S_{\gamma_1}(\mathbf{x})$.

6. Discussions

We point out two directions of future research topics.

On the projected vMF We have shown that the projection score of the vMF distribution onto a geodesic is not exactly vM-distributed, but is so approximately. Whether a similar property holds for more general distributions with ellipse-like symmetry such as Kent distribution (Kent, 1982) and the scaled vM distribution (Scealy and Wood, 2019), or for a semiparametric rotationally symmetric models (see e.g. Paindaveine and Verdebout, 2020), is an interesting question. An anonymous reviewer has asked whether a property similar to Theorem 3.1 holds for semiparametric rotationally symmetric models with density at $\mathbf{x} \in \mathbb{S}^{p-1}$ proportional to $f(\kappa \mathbf{x}'\boldsymbol{\mu})$. A calculation similar to the proof of the theorem shows that if $f(st) = f(s)f(t)$, then the distribution of $T = S_\gamma(\mathbf{x})$ is a scale mixture of $f(\kappa \cos\delta \cos(t))$. Whether the condition on f can be relaxed has not been answered. Furthermore, consider a dimension-k subsphere $A_k = \{\mathbf{P}_k\mathbf{z} : \mathbf{z} \in \mathbb{S}^k\} \subset \mathbb{S}^{p-1}$, defined for a $p \times (k+1)$ orthogonal frame $\mathbf{P}_k$. The subsphere is a multidimensional generalization of the geodesic, as A_1 is the image over a geodesic. A related question is whether the projected score in $\mathbb{S}^k$ onto A_k is approximately vMF-distributed.

On a formal development of projection pursuit for directional data We have demonstrated that departures from vM distributions can be used as useful projection pursuit indices. A comprehensive development of projection pursuit for directional data and a thorough investigation of it are deemed beneficial. We point out two major issues that need to be addressed in the development.

- Projection index: A projection index measures the interestingness of the geodesic projection. While various measures of non-vM-ness are viable candidates, computationally simpler notions such as kurtosis may provide satisfactory exploration of data. As an instance, Alashwali and Kent (2016) primarily used the kurtosis as the projection index in their study of projection pursuit.
- Structure removal or k-dimensional projection pursuit: Our discussion is limited to the dimension-1 projection onto geodesics, and there certainly is a need to detect more than one interesting geodesics or even a subsphere. A common technique to explore more than one interesting directions is structure removal (Friedman, 1987). For directional data, structure removal translates to removing the identified interesting geodesic γ. Such a task is not straightforward and invites a deeper discussion on the strategy of dimension reduction for directional data (and for data on manifolds in general). Roughly, the strategies are categorized into two. The forward dimension reduction aims to successively increase the dimension of "interesting" submanifolds. On the other hand, the *backward* dimension reduction seeks to successively remove "uninteresting" directions. While there are evidences that the backward approaches are more suited for manifold-valued data (Jung, Dryden and Marron, 2012; Huckemann and Eltzner, 2018; Pennec et al., 2018), whether that remains the same for projection pursuit shall be answered. See Damon and Marron (2014) for an introduction on this issue.

Appendix A: Special functions

We collect several facts on the modified Bessel and Struve functions, and their various asymptotic expansions. We refer to Sections 10 and 11 of Olver et al. (2010) and DLMF (2020) as a general reference for these special function.

A.1. Standard power series and integral representations

The modified Bessel function of the first kind of order ν can be defined by the standard power series (Olver et al., 2010, 10.25.2):

$$I_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{2k}}{k!\Gamma(\nu+k+1)}. \tag{A.1}$$

The modified Struve function of the first kind of order ν can be defined by the standard power series (Olver et al., 2010, 11.2.2):

$$L_\nu(z) = (\frac{z}{2})^{\nu+1} \sum_{k=0}^{\infty} \frac{(\frac{z}{2})^{2k}}{\Gamma(k+\frac{3}{2})\Gamma(k+\nu+\frac{3}{2})}. \tag{A.2}$$

The modified Struve function of the second kind of order ν is defined by

$$M_\nu(z) = L_\nu(z) - I_\nu(z), \tag{A.3}$$

and can be expressed as

$$M_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^{k+1} \left(\frac{z}{2}\right)^{\nu+k}}{\Gamma(1+\frac{k}{2})\Gamma(\nu+\frac{k}{2}+1)}. \tag{A.4}$$

These functions, $I_\nu(z), L_\nu(z)$ and $M_\nu(z)$, are defined for $z > 0$ and $\nu > -\frac{1}{2}$. They appear in our discussion as the solution for special forms of integrals. In particular, we have (Olver et al., 2010, 10.32.2, 11.5.4):

$$I_\nu(z) = \frac{(\frac{z}{2})^\nu}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})} \int_{-1}^{1} e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt, \tag{A.5}$$

$$M_\nu(z) = -\frac{2(\frac{z}{2})^\nu}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})} \int_{0}^{1} e^{-zt}(1-t^2)^{\nu-\frac{1}{2}} dt. \tag{A.6}$$

Combining the above with (A.3), we obtain a new integral representation, which is used in the proof of Theorem 3.2.

Lemma A.1. *For $z > 0$ and $\nu > -\frac{1}{2}$,*

$$\int_0^1 e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt = \frac{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}{(\frac{z}{2})^\nu} \left(\frac{1}{2} M_\nu(z) + I_\nu(z) \right).$$

Proof of Lemma A.1. Let $z > 0$ and $\nu > -\frac{1}{2}$ be fixed. Write

$$m_\nu(z) := \frac{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}{(\frac{z}{2})^\nu} M_\nu(z) = -2 \int_0^1 e^{-zt}(1-t^2)^{\nu-\frac{1}{2}} dt$$

and

$$i_\nu(z) := \frac{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}{(\frac{z}{2})^\nu} I_\nu(z) = \int_{-1}^1 e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt.$$

Decomposing the latter integral, we have

$$\begin{aligned} i_\nu(z) &= \int_0^1 e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt + \int_{-1}^0 e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt \\ &= \int_0^1 e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt + \int_0^1 e^{-zt}(1-t^2)^{\nu-\frac{1}{2}} dt \\ &= \int_0^1 e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt - \frac{1}{2} m_\nu(z). \end{aligned}$$

Thus,

$$\int_0^1 e^{zt}(1-t^2)^{\nu-\frac{1}{2}} dt = \frac{1}{2} m_\nu(z) + i_\nu(z). \qquad \square$$

A.2. Large argument asymptotic expressions and related arguments

Various forms of Hankel's expansion can be used to approximate $I_\nu(z)$ for large z. The asymptotic equivalence of $a(z)$ and $b(z)$ is denoted by $a(z) \asymp b(z)$ if $\lim_{z\to\infty} a(z)/b(z) = 1$. As a simple form of Hankel's expansion, as $z \to \infty$ and ν fixed, for any $\ell = 0, 1, \ldots, \infty$,

$$I_\nu(z) \asymp \frac{e^z}{(2\pi z)^{\frac{1}{2}}} \sum_{k=0}^{\ell} (-1)^k \frac{a_k(\nu)}{z^k}, \tag{A.7}$$

where $a_0(\nu) = 1$ and $a_k(\nu) = \frac{(4\nu^2-1^2)(4\nu^2-3^2)\cdots(4\nu^2-(2k-1)^2)}{k!8^k}$ for $k \geq 1$ (Olver et al., 2010, 10.40.1). The remainder term can be controlled as well. For any positive real $z > 0$ (Olver, 1997, p. 269),

$$I_\nu(z) = \frac{e^z}{(2\pi z)^{\frac{1}{2}}} \left\{ \sum_{k=0}^{\ell-1} (-1)^k \frac{a_k(\nu)}{z^k} + R_\ell(-z, \nu) \right\} + O(\frac{e^{-z}}{\sqrt{z}}). \tag{A.8}$$

The remainder term $R_\ell(-z, \nu)$ is absolutely bounded by $|a_\ell(\nu)/z^\ell|$ provided that $\nu < \ell + \frac{1}{2}$ (Nemes, 2017, Eq. (5.1)). While the expression (A.8) is exact for any $z > 0$, it is only useful for large z.

Next two results are on the various asymptotic ratios of the special functions and the exponential function, and are primarily used in high-concentration asymptotic arguments, e.g., in Theorem B.1.

Lemma A.2. *Let $\nu, u \geq 0$ and $\ell > 0$ be fixed.*

(i) $I_\nu(z)/I_u(z) \asymp 1 + \frac{u^2-\nu^2}{2z} + O(z^{-2})$, *as* $z \to \infty$.
(ii) $\lim_{z\to\infty} z^\ell e^{-z} M_\nu(z) = 0$.
(iii) $\lim_{z\to\infty} z^\ell M_\nu(z)/I_u(z) = 0$

Proof of Lemma A.2. (i). Using the first two terms of the Hankel's expansion, we have $I_\nu(z)/I_u(z) \asymp \frac{1-\frac{4\nu^2-1}{8z}+O(z^{-2})}{1-\frac{4u^2-1}{8z}+O(z^{-2})} = 1 - \frac{4\nu^2-1}{8z} + \frac{4u^2-1}{8z} + O(z^{-2})$. The assertion (i) follows.

(ii). Using Dingle's expansion of $M_\nu(z)$ (Dingle, 1973, p. 445) (see also Item 11.6.2 of Olver et al. (2010)), we have

$$\frac{z^\ell}{e^z} M_\nu(z) = \frac{z^\ell}{\pi e^z} \left\{ \sum_{i=0}^{n-1} (-1)^{i+1} \frac{\Gamma(i+\frac{1}{2})(\frac{z}{2})^{\nu-2i-1}}{\Gamma(i-\nu+\frac{1}{2})} + (\frac{z}{2})^{\nu-1-2n} T_n(z) \right\},$$

for any $n > \nu$, and the remainder term $T_n(z)$ satisfies $\sup_{z\in(0,\infty)} T_n(z) < \infty$. Since there is a constant $C = C(\nu) > 0$ such that $\Gamma(i - \nu + \frac{1}{2}) > C$ for all $i = 0, 1, \ldots, n-1$, we have $\lim_{z\to\infty} z^\ell e^{-z} M_\nu(z) = 0$.

Part (iii) is given by the Hankel's expansion of $I_u(z)$ combined with (ii). □

Lemma A.3. *For any given $\nu \geq 0$, the function $z \mapsto M_\nu(z)(\frac{z}{2})^{-\nu}$ is strictly negative and increasing on $z \geq 0$.*

Proof of Lemma A.3. By (A.6), we have

$$\frac{M_\nu(z)}{(\frac{z}{2})^\nu} = -\frac{2}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}\int_0^1 e^{-zt}(1-t^2)^{\nu-\frac{1}{2}}dt,$$

and

$$\frac{d}{dz}\frac{M_\nu(z)}{(\frac{z}{2})^\nu} = -\frac{M_\nu(z)}{(\frac{z}{2})^\nu}.$$

Since $\frac{M_\nu(z)}{(\frac{z}{2})^\nu} < 0$ for any $z \geq 0$, the statement of the lemma follows. □

A.3. Large-order asymptotic expressions and related expansions

The large-order asymptotic expressions for the modified Bessel function of the first kind of order ν, I_ν, and for the modified Struve function of the first kind of order ν, L_ν, are known (Olver et al., 2010, 10.41.1 and 11.6.5): For argument z fixed, as $\nu \to \infty$,

$$I_\nu(z) \asymp \frac{1}{\sqrt{2\pi\nu}}\left(\frac{ez}{2\nu}\right)^\nu, \quad L_\nu(z) \asymp \frac{z}{\pi\nu\sqrt{2}}\left(\frac{ez}{2\nu}\right)^\nu. \tag{A.9}$$

Note that the extra $\sqrt{\nu}$ in the denominator of the approximation of $L_\nu(z)$ makes $L_\nu(z)$ negligible compared to $I_\nu(z)$. Therefore, for the modified Struve function of the second kind M_ν, we have

$$M_\nu(z) = L_\nu(z) - I_\nu(z) \asymp -I_\nu(z), \tag{A.10}$$

and

$$M_\nu(z) + 2I_\nu(z) \asymp I_\nu(z) \asymp \frac{1}{\sqrt{2\pi\nu}}\left(\frac{ez}{2\nu}\right)^\nu. \tag{A.11}$$

Also useful is the the Poincaré expansion of the gamma function (Olver et al., 2010, 5.11.3):

$$\Gamma(\nu) \asymp \sqrt{\frac{2\pi}{\nu}}\left(\frac{\nu}{e}\right)^\nu\left\{1+\frac{1}{12\nu}+O(\nu^{-2})\right\}. \tag{A.12}$$

A.4. Uniform expansion of the modified Bessel function

Uniformly for $0 < z < \infty$,

$$I_\nu(\nu z) = \frac{e^{\nu\eta(z)}}{\sqrt{2\pi\nu}(1+z^2)^{1/4}}\left(1+\frac{1}{8z\nu}+O(\nu^{-2})\right), \tag{A.13}$$

as $\nu \to \infty$ (Olver et al., 2010, 10.41.3), where

$$\eta(z) = \sqrt{1+z^2} + \log z - \log(1+\sqrt{1+z^2}). \tag{A.14}$$

Appendix B: Technical details and proofs

B.1. Proofs of Lemma 2.1, Theorem 3.1, Theorem 3.2, Lemma 3.3 and Theorem 3.4

Proof of Lemma 2.1. We parameterize γ by $\mathbf{p}(t) = \mathrm{Exp}_{\mathbf{q}}(t\mathbf{v}) = \cos(t)\mathbf{q} + \sin(t)\mathbf{v}$ with $t \in (-\pi, \pi]$. Minimizing $\rho(\mathbf{p}(t), \mathbf{x})$ is equivalent to maximizing $\mathbf{x}'\mathbf{p}(t) = \cos(t)\mathbf{x}'\mathbf{q} + \sin(t)\mathbf{x}'\mathbf{v}$. Thus, if $(\mathbf{x}'\mathbf{q}, \mathbf{x}'\mathbf{v}) = (0, 0)$, then the maximum is everywhere. If $(\mathbf{x}'\mathbf{q}, \mathbf{x}'\mathbf{v}) \neq (0, 0)$, then there is a unique polar coordinate $(r, \theta) \in (0, 1] \times (-\pi, \pi]$ of $(\mathbf{x}'\mathbf{q}, \mathbf{x}'\mathbf{v})$, so that $\mathbf{x}'\mathbf{p}(t) = r\cos(t - \theta)$. Thus $S_\gamma(\mathbf{x}) = \mathrm{argmax}_t\, \mathbf{x}'\mathbf{p}(t) = \theta$. □

Proof of Theorem 3.1. In this proof, the notation $f(x_1, \cdots, x_m)$ is generically used for the (joint) density function of variables $(x_1, \ldots, x_m)$. Without losing generality, let $\boldsymbol{\mu} = \mathbf{e}_1$, $\mathbf{q} = \cos(\delta)\mathbf{e}_1 + \sin(\delta)\mathbf{e}_2$ and $\mathbf{v} = \mathbf{e}_3$.

Write $\mathbf{x} = (x_1, \ldots, x_p)'$ for $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$. In the spherical coordinate system $(\theta_1, \ldots, \theta_{p-2}, \phi)$, where $\theta_i \in [0, \pi)$ $(i = 1, \ldots, p-2)$, $\phi \in [-\pi, \pi)$, and

$$x_1 = \cos\theta_1,\ x_2 = \sin\theta_1\cos\theta_2, \ldots,\ x_p = \left(\prod_{i=1}^{p-2}\sin\theta_i\right)\sin\phi,$$

we have $f(\theta_1, \ldots, \theta_{p-2}, \phi) = c_p(\kappa)\exp(\kappa\cos\theta_1)\prod_{j=1}^{p-2}(\sin\theta_j)^{p-j-1}$.

Suppose that $p \geq 5$. Then the marginal density function of $(\theta_1, \theta_2, \theta_3)$ is

$$f(\theta_1, \theta_2, \theta_3) = c_p(\kappa)S_{p-4}\exp(\kappa\cos\theta_1)(\sin\theta_1)^{p-2}(\sin\theta_2)^{p-3}(\sin\theta_3)^{p-4}, \quad \text{(B.1)}$$

for $\theta_i \in [0, \pi)$. Here,

$$S_w = \frac{2\pi^{\frac{w+1}{2}}}{\Gamma(\frac{w+1}{2})},$$

for $w \geq 0$ is surface area of the w-dimensional unit sphere in dimension $w + 1$. From (B.1), the joint density function of (x_1, x_2, x_3) is obtained as

$$f(x_1, x_2, x_3) = c_p(\kappa)S_{p-4}e^{\kappa x_1}(1 - x_1^2 - x_2^2 - x_3^2)^{\frac{p-5}{2}}, \quad \text{(B.2)}$$

for x_1, x_2, x_3 satisfying $x_1^2 + x_2^2 + x_3^2 \leq 1$. A similar derivation for the $p = 4$ case, involving ϕ, shows that (B.2) holds for $p = 4$ as well.

Let $z_1 = x_1\cos\delta + x_2\sin\delta$, $z_2 = -x_1\sin\delta + x_2\cos\delta$, and $z_3 = x_3$. From $f(z_1, z_2, z_3)$, the joint density of $(\mathbf{x}'\mathbf{q}, \mathbf{x}'\mathbf{v}) = (z_1, z_3)$ is obtained by marginalizing out z_2. This is done by utilizing the integral representation of the modified Bessel function I_ν in (A.5). Also using $S_{p-4} = S_{p-3}\Gamma(\nu)/\{\sqrt{\pi}\Gamma(\nu - \frac{1}{2})\}$, we get

$$f(z_1, z_3) = c_p(\kappa)S_{p-3}e^{\kappa\cos(\delta)z_1}(1 - r^2)^{\nu-1}I^*_{\nu-1}(z)\Gamma(\nu) \quad \text{(B.3)}$$

for $z_1^2 + z_3^2 \leq 1$, where $\nu = (p-2)/2$, $r^2 = z_1^2 + z_3^2$ and $z = \kappa\sin(\delta)\sqrt{1 - r^2}$ and $I^*_\nu(z) = (\frac{z}{2})^{-\nu}I_\nu(z)$.

For $p = 3$, calculations show $f(z_1, z_3) = c_3(\kappa)e^{\kappa\cos(\delta)z_1}(1-r^2)^{-\frac{1}{2}}2\cosh(z)$. Using the relation $I_{-\frac{1}{2}}(z) = (\frac{2}{\pi z})^{\frac{1}{2}}\cosh z$ (Olver et al., 2010, 10.39.1) and $S_0 = 2$ confirms that (B.3) holds for the $p = 3$ case as well. Switching to the polar coordinates (r, t) where $z_1 = r\cos t, z_3 = r\sin t$,

$$f(r,t) = c_p(\kappa)S_{p-3}\exp\{\kappa\cos(\delta)r\cos(t)\}r(1-r^2)^{\nu-1}I^*_{\nu-1}(z)\Gamma(\nu). \tag{B.4}$$

Using $\int_{-\pi}^{\pi} e^{z\cos t}dt = 2\pi I_0(z)$, the marginal density of r is easily obtained. Plugging in the expression of $c_p(\kappa)$ given in (1.1) and $S_{p-3} = (2\pi^\nu)/\Gamma(\nu)$ proves the assertion. □

Proof of Theorem 3.2. The joint density function of (R, T), where $R = S_\gamma(\mathbf{x})$, $R\cos T = \mathbf{q}'\mathbf{x}$ and $R\sin T = \mathbf{v}'\mathbf{x}$, is given by Theorem 3.1, or more directly by equation (B.4) in the proof of Theorem 3.1. Setting $\delta = 0$, we have

$$f(r,t) = c_p(\kappa)S_{p-3}e^{(\kappa\cos t)r}r(1-r^2)^{\frac{p-4}{2}}. \tag{B.5}$$

Due to Lemma 2.1, the distribution of the projection score $t = S_\gamma(\mathbf{x})$ is obtained by marginalizing (B.5). Integration by parts gives the marginal density of t, $f(t) = c_p(\kappa)S_{p-3}\left(\frac{1}{p-2} + g(t)\right)$, where

$$g(t) = \frac{\kappa\cos t}{p-2}\int_0^1 e^{(\kappa\cos t)r}(1-r^2)^{\frac{p-2}{2}}dr. \tag{B.6}$$

If $t = \pm\pi/2$, then $g(t) = 0$. Thus,

$$f(\pi/2) = c_p(\kappa)S_{p-3}\frac{1}{p-2}.$$

In general, there is no closed-from expression of $g(t)$. If $|t| \in (\pi/2, \pi]$, that is, if $\kappa\cos t < 0$, then the integral in (B.6) is closely related to the modified Struve function M_ν. Using the integral representation of M_ν in (A.6),

$$\begin{aligned} g(t) &= \frac{\kappa\cos t}{p-2}\int_0^1 e^{(\kappa\cos t)r}(1-r^2)^{\frac{p-2}{2}}dr \\ &= \frac{-\kappa\cos t}{p-2}\frac{\sqrt{\pi}\Gamma(\frac{p}{2})}{2\left(\frac{-\kappa\cos t}{2}\right)^{\frac{p-1}{2}}}M_\nu(-\kappa\cos t). \end{aligned}$$

Suppose now $|t| < \pi/2$ so that $\kappa\cos t > 0$. It is shown in Lemma A.1 that the integral (B.6) is a linear combination of $M_\nu(\kappa\cos t)$ and $I_\nu(\kappa\cos t)$:

$$g(t) = \frac{\kappa\cos t}{p-2}\frac{\sqrt{\pi}\Gamma(\frac{p}{2})}{\left(\frac{\kappa\cos t}{2}\right)^{\frac{p-1}{2}}}\left(\frac{1}{2}M_\nu(\kappa\cos t) + I_\nu(\kappa\cos t)\right).$$

The derivation is completed by combining above results with the definitions of $c_p(\kappa)$ and S_{p-3}. □

Proof of Lemma 3.3. Without loss of generality, let $\boldsymbol{\mu} = \mathbf{e}_1$, $\mathbf{q} = \mathbf{e}_{p-1}$ and $\mathbf{v} = \mathbf{e}_p$. Then in the spherical coordinate system used in the proof of Theorem 3.2,

$\mathbf{q}'\mathbf{x} = \prod_{j=1}^{p-2} \sin(\theta_j)\cos(\phi)$ and $\mathbf{v}'\mathbf{x} = \prod_{j=1}^{p-2} \sin(\theta_j)\sin(\phi)$, which in turn leads that $\phi = S_\gamma(\mathbf{x})$. With the given spherical coordinates, the marginal density of ϕ is uniform for any rotationally symmetric distribution about $\mathbf{e}_1$ on $\mathbb{S}^{p-1}$, as shown in Lemma 3 of Kim, Schulz and Jung (2020). □

Proof of Theorem 3.4. The Jones–Pewsey distribution has the unique mode at μ. Since f_{PvM} has the mode at 0, we do not consider Jones–Pewsey distributions whose mode is not at 0.

We will prove the theorem by contradiction. Fix any $\kappa \in (0, \infty)$. Suppose that there exists $(\tilde{\kappa}, \phi)$ such that $g_{\mathrm{JP}}(t; \tilde{\kappa}, \phi)$ is equal to $f_{\mathrm{PvM}}(t; p, \kappa)$ for all $t \in I$, where $I \subset (-\pi/2, \pi/2)$ is an open set. Let $z = z(t) = \cos t$. By expressing both density functions as power series in terms of $\cos t$, we have for $g^*(z) := g_{\mathrm{JP}}(t; \tilde{\kappa}, \phi)$,

$$g^*(z) = \omega(\tilde{\kappa}, \phi)\left\{1 + \frac{\tau}{\phi}z - \frac{\tau^2(\phi-1)}{2\phi^2}z^2 + \frac{\tau^3(\phi-1)(2\phi-1)}{6\phi^3}z^3 + \cdots\right\},$$

where $\tau = \tanh(\tilde{\kappa}\phi)$. Using $M_u(z) + 2I_u(z) = L_u(z) + I_u(z)$, given by (A.3), we have for $z > 0$,

$$\begin{aligned} f^*(z) := f_{\mathrm{PvM}}(t; p, \kappa) &= C(p, \kappa)\left\{1 + \frac{\Gamma(\frac{1}{2})\Gamma(u+\frac{1}{2})}{(\frac{\kappa z}{2})^{u-1}}(L_u(\kappa z) + I_u(\kappa z))\right\} \\ &= C(p, \kappa)\left\{1 + \frac{B(\frac{1}{2}, u+\frac{1}{2})\kappa}{2}z + \frac{\kappa^2}{2(u+\frac{1}{2})}z^2 + \frac{B(\frac{1}{2}, u+\frac{1}{2})\kappa^3}{8(u+1)}z^3 + \cdots\right\}, \end{aligned}$$

where $u = (p-1)/2$. We used the standard power series representations for the modified Bessel and Struve functions in (A.1) and (A.2).

Since by assumption $g^*(z) = f^*(z)$ for all $z \in \cos(I)$, where $\cos(I)$ is an open set in $(0, 1)$, the coefficients of both power series should be equal. By matching the first four coefficients, we have $C(p, \kappa) = \omega(\tilde{\kappa}, \phi)$, $\tau/\phi = B(\frac{1}{2}, u+\frac{1}{2})\frac{\kappa}{2}$, $\tau^2(\phi-1)/(2\phi^2) = -\kappa^2/\{2(u+\frac{1}{2})\}$, and $\{\tau^3(\phi-1)(2\phi-1)\}/(6\phi^3) = \{B(\frac{1}{2}, u+\frac{1}{2})\kappa^3\}/\{8(u+1)\}$. Comparing the latter three, we must have

$$\phi = 1 - \frac{4}{B^2(\frac{1}{2}, u+\frac{1}{2})(u+\frac{1}{2})} = \frac{1}{2}\left\{1 - \frac{3(u+\frac{1}{2})}{2(u+1)}\right\},$$

which does not hold for any $u > 0$. Therefore, for any $p \geq 3$, there is no ϕ for any $\tilde{\kappa}$ for which $g_{\mathrm{JP}}(t; \tilde{\kappa}, \phi)$ equals $f_{\mathrm{PvM}}(t; p, \kappa)$. □

B.2. Proof of Theorem 4.1

We first give a sketch of the proof of Theorem 4.1, then supply a detailed proof.

Sketch of the proof of Theorem 4.1. As the exact expression of the density $f_{\mathrm{PvM}}(\cdot; p, \kappa)$ consists of $I_u(z)$, $M_u(z)$ and the gamma function, inspections of these functions are necessary.

For high concentration case, the argument z of $I_u(z)$ and $M_u(z)$ are increasing. We use Hankel's large-argument expansion of the modified Bessel function to approximate $I_u(z)$ as $z \to \infty$; see (A.7) and (A.8). We use the first two terms of the Hankel's expansion while providing a bound for the remainder. In Lemma A.2, the modified Struve function $M_u(z)$ is shown to be asymptotically negligible when compared to the exponential function e^z or to $I_u(z)$ for large z. Then, $|f_\kappa(s;p,\kappa) - g_\kappa(s;\kappa)|$ is evaluated using the expansions; see Theorem B.1. In addition, we give an independent proof for the approximation of the vM by the normal distribution in high concentration in Lemma B.2. The assertion (i) is given by the above two.

The low concentration case (ii) is handled by the standard power series representations of $I_u(z)$ and $M_u(z)$, given in (A.1)–(A.4). We show $f_{\mathrm{PvM}}(t;p,\kappa) = f_{\mathrm{vM}}(t;\tilde{\kappa}_l) + \Theta(\kappa^2)$ as $\kappa \to 0$ in Theorem B.4. An independent proof for the approximation of the vM by Cardioid distribution in low concentration, $f_{\mathrm{vM}}(t;\kappa) = f_{\mathrm{Cardioid}}(t;\kappa/2) + \Theta(\kappa^2)$, is given in Lemma B.3.

For the high-dimension case (iii), the order u of $I_u(z)$ and $M_u(z)$ increases. The large-order asymptotic expressions in (A.9) and (A.11) and the Poincaré expansion of the gamma function (A.12) are used to show $f_{\mathrm{PvM}}(t;p,\kappa) = f_{\mathrm{Cardioid}}(t;\tilde{\kappa}_d/2) + \Theta(p^{-1})$ as $p \to \infty$. Applying Lemma B.3 completes the proof of (iii).

Finally, for the high-concentration, high-dimensional case (iv), the uniform expansion of the modified Bessel function of the first kind (A.13) is useful: Uniformly for $0 < z < \infty$,

$$I_\nu(\nu z) = \frac{e^{\nu\eta(z)}}{\sqrt{2\pi\nu}(1+z^2)^{1/4}}\left(1 + \frac{1}{8z\nu} + O(\nu^{-2})\right),$$

as $\nu \to \infty$. For example, the uniform expansion can be applied to the expression $I_u(\kappa) = I_u(u\lambda)$, as both $u = (p-1)/2$ and κ increase but $\lambda = \kappa/u$ is fixed. In our application, the uniform expansion is carefully applied to expand $I_{u-\frac{1}{2}}(uz)$ as $u \to \infty$. It turns out $I_{u-\frac{1}{2}}(uz)$ is not asymptotically equivalent to $I_u(uz)$. These issues are taken care of in Theorem B.6 where we show $f_\kappa(s;p,\kappa) = g_\kappa(s;\tilde{\kappa}_b) + \Theta(p^{-1})$. The proof for (iv) is completed by applying Lemma B.2 together with a change of variable. □

In addition to the big Theta notation, we use the following. For asymptotic equivalence we say $a(\nu) \asymp b(\nu)$ if $\lim_{\nu\to\infty} a(\nu)/b(\nu) = 1$. We also use the standard big O notation: $f_\nu = O(g_\nu)$ if $\limsup_{\nu\to\infty} |f_\nu|/g_\nu < \infty$. Note that $f_\nu \asymp g_\nu$ implies $f_\nu = \Theta(g_\nu)$, which in turn implies $f_\nu = O(g_\nu)$. The inverses are not in general true.

B.2.1. High concentration approximation

When the concentration parameter κ in $\mathrm{vMF}_p(\boldsymbol{\mu},\kappa)$ is large, the vMF density is very similar to the spherical normal distribution with variance $1/\kappa$ (Kent, 1978). Similar to this result, the geodesic-projected vMF density f_{PvM} can be approximated by a vM density.

As $\kappa \to \infty$, both $\mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$ and $f_{\mathrm{PvM}}(p, \kappa)$ degenerates to their respective mode. In the following, we compare the density of $\sqrt{\kappa} S_\gamma(\mathbf{x})$ with that of $\sqrt{\kappa}\theta$, for $\theta \sim \mathrm{vM}(\kappa)$. Denoting the density functions of $\sqrt{\kappa} S_\gamma(\mathbf{x})$ and $\sqrt{\kappa}\theta$ by $f_\kappa(s)$ and $g_\kappa(s)$, respectively, we derive the following result on the similarity of $f_\kappa(s)$ to $g_\kappa(s)$.

Theorem B.1. *For all $p \geq 3$, and for any $s \in \mathbb{R}$,*

$$f_\kappa(s) = g_\kappa(s) + \Theta(\kappa^{-1}),$$

as $\kappa \to \infty$.

Proof of Theorem B.1. The scaled PvM density is $f_\kappa(s) = C(p, \kappa)/\sqrt{\kappa} \cdot (1 + h(s/\sqrt{\kappa}))$, where the definition of $h(t)$ is given in Theorem 3.2. The scaled vM density is decomposed into $g_\kappa(s) = g_1(\kappa) \cdot g_2(s; \kappa)$, where, for $u = (p-1)/2$,

$$g_1(\kappa) = \frac{1}{2\sqrt{\kappa}\pi I_0(\kappa)} \frac{(\frac{\kappa}{2})^{u-\frac{1}{2}}}{\Gamma(u + \frac{1}{2})}, \quad g_2(s; \kappa) = \frac{\Gamma(u + \frac{1}{2})}{(\frac{\kappa}{2})^{u-\frac{1}{2}}} e^{\kappa \cos(s/\sqrt{\kappa})}.$$

Using the above factorization of density functions, $|f_\kappa(s) - g_\kappa(s)| \leq A(s; \kappa) + B(s; \kappa)$, where

$$\begin{aligned} A(s; \kappa) &= |C(p, \kappa)/\{\sqrt{\kappa} g_1(\kappa)\} - 1| \cdot g_1(\kappa) |1 + h(s/\sqrt{\kappa})|, \\ B(s; \kappa) &= g_1(\kappa) |1 + h(s/\sqrt{\kappa}) - g_2(s; \kappa)|. \end{aligned}$$

It is sufficient to show that $A(s; \kappa) = O(\kappa^{-1})$ and $B(s; \kappa) = \Theta(\kappa^{-1})$ as $\kappa \to \infty$. We record the following for multiple use. Applying Hankel's large κ approximation of $I_0(\kappa)$ given in (A.7), we have, as $\kappa \to \infty$,

$$g_1(\kappa) = \Theta(\kappa^{u-\frac{1}{2}} e^{-\kappa}). \tag{B.7}$$

(i) A bound on $A(s; \kappa)$. Since f_{PvM} is symmetric and has the unique mode at 0, the maximum of $1 + h(s/\sqrt{\kappa})$ is attained at $s = 0$. That is, $|1 + h(s/\sqrt{\kappa})| \leq 1 + h(0)$. To analyze

$$g_1(\kappa)\{1 + h(0)\} = g_1(\kappa) + \frac{1}{\sqrt{2\pi}} \frac{M_u(\kappa)}{I_0(\kappa)} + \frac{2}{\sqrt{2\pi}} \frac{I_u(\kappa)}{I_0(\kappa)},$$

we use (B.7) and the asymptotic expressions of the ratios of the modified Bessel and Struve functions, derived in Lemma A.2 (i) and (iii). We get:

$$g_1(\kappa) \sup_s |1 + h(s/\sqrt{\kappa})| \asymp \frac{2}{\sqrt{2\pi}} \left\{ 1 - \frac{u^2}{2\kappa} + O(\kappa^{-2}) \right\}. \tag{B.8}$$

We also observe that, by Lemma A.2 (i),

$$\frac{\kappa^{-\frac{1}{2}} C(p, \kappa)}{g_1(\kappa)} - 1 = \frac{I_0(\kappa)}{I_\nu(\kappa)} - 1 \asymp \frac{\nu^2}{2\kappa} + O(\kappa^{-2}), \tag{B.9}$$

where $\nu = (p-1)/2$. Combining (B.8) and (B.9) shows that for any $s \in \mathbb{R}$, $A(s;\kappa) = O(\kappa^{-1})$ as $\kappa \to \infty$.

(ii) A bound on $B(s;\kappa)$. Choose κ large enough to satisfy $\kappa > (2s/\pi)^2$ so that $|s|/\sqrt{\kappa} < \pi/2$. Then, $h(s/\sqrt{\kappa})$ has two terms (see the first line of (3.3)), and we write $h(s/\sqrt{\kappa}) = h_M(x) + h_I(x)$, where $x = \cos(s/\sqrt{\kappa}) \in (0,1]$,

$$h_M(x) = \frac{\sqrt{\pi}\Gamma(u+\frac{1}{2})}{(\frac{\kappa x}{2})^{u-1}} M_u(\kappa x), \quad h_I(x) = \frac{\sqrt{\pi}\Gamma(u+\frac{1}{2})}{(\frac{\kappa x}{2})^{u-1}} 2I_u(\kappa x).$$

Define g_I to be the function satisfying $g_I(x) = g_2(s;\kappa)$ for $x = \cos(s/\sqrt{\kappa})$. Then,

$$\begin{aligned} &|1 + h(s/\sqrt{\kappa}) - g_2(s;\kappa)| \\ &\quad = |1 + h_M(x) + h_I(x) - g_I(x)| \\ &\quad \le \sup_{x\in(0,1]} |1 + h_M(x)| + |h_I(x) - g_I(x)|. \end{aligned} \tag{B.10}$$

The first term of (B.10) is bounded:

$$\begin{aligned} \sup_{x\in(0,1]} |1 + h_M(x;\kappa)| &\le 1 + \sqrt{\pi}\Gamma\left(u+\frac{1}{2}\right) \sup_{x\in(0,1)} \left|\frac{M_u(kx)}{(\frac{\kappa x}{2})^u}\right| \frac{\kappa}{2} \\ &= 1 + \frac{\Gamma(\frac{1}{2})\Gamma(u+\frac{1}{2})}{\Gamma(u+1)} \frac{\kappa}{2}. \end{aligned} \tag{B.11}$$

Here, we used the fact that $\frac{M_u(z)}{(\frac{z}{2})^u}$ is strictly negative and increasing on $z \ge 0$, shown in Lemma A.3. Thus,

$$\sup_{x\in(0,1]} \left|\frac{M_u(kx)}{(\frac{\kappa x}{2})^u}\right| = \lim_{x\to 0} \left|\frac{M_u(kx)}{(\frac{\kappa x}{2})^u}\right| = \frac{1}{\Gamma(u+1)}. \tag{B.12}$$

Note that the limit in (B.12) is easily obtained by replacing $M_u(z)$ with its power series representation (A.4).

For the second term of (B.10), we use the exact Hankel's expansion of $I_1(\kappa)$ in (A.8) and write

$$\begin{aligned} |h_I(x) - g_I(x)| &= \frac{\Gamma(u+\frac{1}{2})}{(\frac{\kappa}{2})^{u-\frac{1}{2}}} |\sqrt{2\pi\kappa} I_u(\kappa x) - \exp(\kappa x)| \\ &= \frac{\Gamma(u+\frac{1}{2})}{(\frac{\kappa}{2})^{u-\frac{1}{2}}} \left| \sqrt{2\pi\kappa} \frac{e^{\kappa x}}{\sqrt{2\pi\kappa x}} \left\{ 1 + \frac{4u^2-1}{8\kappa x} + O(\kappa^{-2}) \right\} - \exp(\kappa x) \right| \\ &= \frac{\Gamma(u+\frac{1}{2})}{(\frac{\kappa}{2})^{u-\frac{1}{2}}} \frac{e^{\kappa x}}{\sqrt{x}} \left(1 - \sqrt{x} + \frac{4u^2-1}{8\kappa} + O(\kappa^{-2}) \right) \\ &= \Theta(\kappa^{-u-\frac{1}{2}} e^{\kappa}). \end{aligned} \tag{B.13}$$

In the last asymptotic equality, we used $x = \cos(s/\sqrt{\kappa}) \to 1$ as $\kappa \to \infty$.

Combining (B.7), (B.10), (B.11) and (B.13) gives

$$\begin{aligned} B(s;\kappa) &= g_1(\kappa)|1 + h(\kappa^{-\frac{1}{2}}s) - g_2(s;\kappa)| \\ &= \Theta\left\{(\kappa^{u-\frac{1}{2}}e^{-\kappa})\kappa^{-u-\frac{1}{2}}e^{\kappa}\right\} = \Theta(\kappa^{-1}), \end{aligned}$$

as desired. □

The next lemma is on the approximation of vM distributions by the normal distribution when $\kappa \to \infty$. Let $\phi(s)$ be the density function of the standard normal distribution.

Lemma B.2. *For any $s \in \mathbb{R}$, as $\kappa \to \infty$,*

$$g_\kappa(s) = \phi(s) + \Theta(\kappa^{-1}).$$

Proof of Lemma B.2. Using the Hankel's expansion (A.8), and taking the power series representation of $\cos(t)$ and $\exp(z)$, we have

$$\begin{aligned} g_\kappa(s) - \phi(s) &= \frac{e^{\kappa\cos(s/\sqrt{\kappa})}}{2\pi I_0(\kappa)} - \frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}} \\ &= \frac{e^{-\frac{s^2}{2}}}{\sqrt{2\pi}}\left[\frac{\exp\{\kappa(1-\frac{s^2}{2\kappa}+\frac{s^4}{4!\kappa^2}+O(\kappa^{-3}))-\frac{s^2}{2}\}}{e^{\kappa}\left(1-\frac{1}{8\kappa}+O(\kappa^{-2})\right)} - 1\right] \\ &= \frac{e^{-\frac{s^2}{2}}}{\sqrt{2\pi}}\left\{\left(\frac{1}{8}+\frac{s^4}{4!}\right)\frac{1}{\kappa}+O(\kappa^{-2})\right\}. \end{aligned}$$

□

B.2.2. Low concentration approximation

We first give the statement and an independent proof on the similarity of the vM distribution and Cardioid distribution when κ is small.

Lemma B.3. *For any $t \in (-\pi, \pi)$, as $\kappa \to 0$,*

$$f_{\mathrm{vM}}(t;\kappa) = f_{\mathrm{Cardioid}}(t;\kappa/2) + \Theta(\kappa^2).$$

Proof of Lemma B.3. By using the standard power series representation for the modified Bessel function (A.1), we write

$$\begin{aligned} f_{\mathrm{vM}}(t;\kappa) - f_{\mathrm{Cardioid}}(t;\kappa/2) &= \frac{1}{2\pi}\left\{\frac{e^{\kappa\cos t}}{I_0(\kappa)} - (1+\kappa\cos t)\right\} \\ &= \frac{1}{2\pi}\left\{\frac{1+\kappa\cos t+\frac{(\kappa\cos t)^2}{2}+O(\kappa^3)}{1+(\frac{\kappa}{2})^2+O(\kappa^4)} - (1+\kappa\cos t)\right\} \\ &= \frac{1}{2\pi}\left\{\frac{2\cos^2 t-1}{4}\kappa^2+O(\kappa^3)\right\} = \Theta(\kappa^2). \end{aligned}$$

□

Our main result for the low-concentration case follows:

Theorem B.4. *Let* $\tilde{\kappa} = \tilde{\kappa}(\kappa) = B(\frac{p}{2}, \frac{1}{2})\kappa/2$. *For all* $p \geq 3$, *and for any* $t \in (-\pi, \pi]$, *as* $\kappa \to 0$,

$$f_{\mathrm{PvM}}(t; p, \kappa) = f_{\mathrm{vM}}(t; \tilde{\kappa}) + \Theta(\kappa^2).$$

Proof of Theorem B.4. Recall that $f_{\mathrm{vM}}(t; \tilde{\kappa}) = c_2(\tilde{\kappa}) \exp(\tilde{\kappa} \cos t)$, where $c_2(\tilde{\kappa}) = \{2\pi I_0(\tilde{\kappa})\}^{-1}$. Write

$$\begin{aligned} |f_{\mathrm{PvM}}(t; p, \kappa) - f_{\mathrm{vM}}(t; \tilde{\kappa})| &\leq |C(p, \kappa) - c_2(\tilde{\kappa})|(1 + h(t)) \\ &\quad + c_2(\tilde{\kappa})|1 + h(t) - \exp(\tilde{\kappa} \cos t)|. \end{aligned} \tag{B.14}$$

By using the standard power series representation for the modified Bessel function in (A.1), we get $\lim_{\kappa \to 0} c_2(\tilde{\kappa}) = 1/(2\pi)$ and $|C(p, \kappa)/c_2(\tilde{\kappa}) - 1| = |\{1 + \tilde{\kappa}^2/4 + O(\kappa^4)\} - 1|$, leading to

$$C(p, \kappa) - c_2(\tilde{\kappa}) \asymp \frac{1}{8\pi}\tilde{\kappa}^2 + O(\kappa^4). \tag{B.15}$$

Since f_{PvM} is symmetric and has the unique mode at 0,

$$1 \leq 1 + h(t) \leq 1 + h(0), \tag{B.16}$$

which is bounded below and above uniformly for $\kappa \in (0, 1]$. (The case $\kappa > 1$ is not considered as $\kappa \to 0$.)

For evaluation of $m(t) := 1 + h(t) - \exp(\tilde{\kappa} \cos t)$, we inspect three separate cases depending on the sign of $\cos t$. First, if $\cos t = 0$, $m(t) = 0$. Second, if $\cos t > 0$, then using (A.1) and (A.2),

$$\begin{aligned} 1 + h(t) &= \sum_{j=1}^{\infty} \frac{\Gamma(\frac{1}{2})\Gamma(\frac{p}{2})}{\Gamma(\frac{1}{2} + \frac{j}{2})\Gamma(\frac{p}{2} + \frac{j}{2})} \left(\frac{\kappa \cos t}{2}\right)^j \\ &= 1 + B(\tfrac{p}{2}, \tfrac{1}{2})\frac{\kappa \cos t}{2} + \frac{\Gamma(\frac{1}{2})\Gamma(\frac{p}{2})}{\Gamma(\frac{1}{2} + 1)\Gamma(\frac{p}{2} + 1)} \left(\frac{\kappa \cos t}{2}\right)^2 + O(\kappa^3), \end{aligned}$$

and

$$\begin{aligned} m(t) &= 1 + h(t) - \sum_{j=0}^{\infty} \frac{(\tilde{\kappa} \cos t)^j}{j!} \\ &= \left[\frac{4}{p} - \frac{1}{2}\{B(\tfrac{p}{2}, \tfrac{1}{2})\}^2\right] \left(\frac{\kappa \cos t}{2}\right)^2 + O(\kappa^3). \end{aligned} \tag{B.17}$$

For the case $\cos t < 0$, a similar derivation as above using the power series representation for $M_u(z)$ in (A.4) shows that $m(t)$ is of order κ^2 as well.

The proof is completed by combining (B.14)–(B.17). □

B.2.3. High dimensional approximation

Theorem B.5. *For any given $\kappa > 0$ and $t \in (-\pi, \pi]$, as $p \to \infty$,*

$$f_{\mathrm{PvM}}(t; p, \kappa) = f_{\mathrm{Cardioid}}(t; \rho) + \Theta(p^{-1}),$$

where $\rho = \frac{1}{2}\sqrt{\frac{\pi}{2(p-1)}}\kappa$.

Proof of Theorem B.5. We analyze

$$\begin{aligned} f_{\mathrm{PvM}}(t; p, \kappa) - f_{\mathrm{Cardioid}}(t; \rho) &= C(p, \kappa)\{1 + h(t)\} - \frac{1}{2\pi}(1 + 2\rho\cos t) \\ &= \frac{1}{2\pi}\left[\{1 - 2\pi C(p, \kappa)\}\{1 + h(t)\} + \{h(t) - 2\rho\cos t\}\right]. \end{aligned}$$

Utilizing the Poincaré expansion of the gamma function (A.12) and the large-order asymptotic expression for I_u in (A.9), we get $1 - 2\pi C(p, \kappa) \asymp -\frac{1}{6(p-2)} + O(p^{-2})$. Thus, it is sufficient to show that $h(t) \to 0$ and $h(t) - 2\rho\cos t = \Theta(p^{-\frac{3}{2}})$ as $p \to \infty$. Since the latter implies the former, it remains to show $h(t) - 2\rho\cos t = \Theta(p^{-\frac{3}{2}})$.

Let $u = (p-1)/2$. For $|t| > \pi/2$, $h(t)$ is expanded using (A.12) and the large-order asymptotic expression for M_u in (A.9)–(A.10). We have for $u \to \infty$,

$$\begin{aligned} h(t) &= \frac{\sqrt{\pi}\Gamma(u + \frac{1}{2})}{\left(\frac{|\kappa\cos t|}{2}\right)^{u-1}} M_u(|\kappa\cos t|) \\ &\asymp -\sqrt{\pi}\left(1 + \frac{1}{2u}\right)^u e^{-\frac{1}{2}} \frac{|\kappa\cos t|}{2\sqrt{u}} \left\{1 + \frac{1}{12u} + O(\nu^{-2})\right\} \\ &\asymp \frac{\sqrt{\pi}\kappa}{2\sqrt{u}}\cos t + \frac{\sqrt{\pi}\kappa\cos t}{24} u^{-\frac{3}{2}} + O(u^{-5/2}). \end{aligned} \tag{B.18}$$

In (B.18), we used the fact $\lim_{u\to\infty}(1 + \frac{1}{2u})^u = e^{1/2}$ and that $\cos t < 0$ for $|t| \in (\pi/2, \pi)$. For $|t| < \pi/2$, a similar derivation done for (B.18), but using (A.9) and (A.11), gives the same asymptotic result. Thus, for $t \neq \pm\pi/2$, $h(t) - 2\rho\cos t = \Theta(p^{-\frac{3}{2}})$. Observing that $h(t) = 2\rho\cos t = 0$ for $t = \pm\pi/2$ completes the proof. □

B.2.4. Approximation along high dimension and high concentration

We consider $\kappa = \lambda u$, where $u = (p-1)/2$, for a fixed $\lambda > 0$ and $u \to \infty$. Recall that for $T \sim \mathrm{PvM}(p, \kappa)$ and $\theta \sim \mathrm{vM}(w)$, we denote the density functions of $\sqrt{\kappa}S_\gamma(\mathbf{x})$ and $\sqrt{\kappa}\theta$ by $f_\kappa(s)$ and $g_\kappa(s)$, respectively. Here, $\kappa > 0$ and $w > 0$ need not be the same.

Theorem B.6. *Let $u = (p-1)/2$ and $\kappa = \lambda u$ for a fixed $\lambda \in (0, \infty)$. Let $w = \frac{\lambda^2 u}{1+\sqrt{1+\lambda^2}}$. Then, for any $s \in \mathbb{R}$,*

$$f_\kappa(s; p, \kappa) = g_\kappa(s; w) + \Theta(u^{-1})$$

as $u \to \infty$.

Remark 3. The statement of Theorem B.6 holds when $u = (p-1)/2$ is replaced by $\nu = (p-2)/2$ or $p/2$.

Proof of Theorem B.6. Note that $f_\kappa(s) = \kappa^{-1/2}C(p,\kappa)\{1 + h(s/\sqrt{\kappa})\}$ and $g_k(s) = \kappa^{-1/2}\{2\pi I_0(w)\}^{-1}\exp\{w\cos(s/\sqrt{\kappa})\}$. Let

$$\nu = (p-2)/2 = u - \frac{1}{2}.$$

As $u \to \infty$, $\kappa = \lambda u \to \infty$ and $\nu \to \infty$ as well.

We first show that $\kappa^{-1/2}C(p,\kappa)$ degenerates to 0 in the limit. Using the Poincaré expansion of the gamma function (A.12) and the uniform expansion of the modified Bessel function (A.13), we obtain

$$\kappa^{-1/2}C(p,\kappa) \asymp \frac{1}{2\pi}\left(\frac{\sqrt{1+\lambda^2}}{\lambda}\frac{e}{\nu}\right)^{\frac{1}{2}}\exp\{-\nu\xi(\lambda,\nu)\},$$

where $e = \exp(1)$ and using $a_\nu = (1 + \frac{1}{2\nu})^2$,

$$\xi(\lambda,\nu) = \sqrt{1+a_\nu\lambda^2} + \log a_\nu - \log\left(1+\sqrt{1+a_\nu\lambda^2}\right) - 1 + \log 2.$$

Checking $\xi(0,\nu) = \log(2) > 0$ and $\frac{\partial}{\partial\lambda}\xi(\lambda,\nu) = \frac{a_\nu\lambda}{1+\sqrt{1+a_\nu\lambda^2}} > 0$ confirms that $\xi(\lambda,\nu) > 0$ for any λ and ν. In particular,

$$\lim_{\nu\to\infty}\xi(\lambda,\nu) = \sqrt{1+\lambda^2} - 1 + \log 2 - \log(1+\sqrt{1+\lambda^2}) > 0$$

for any $\lambda > 0$. As $\nu \to \infty$,

$$\kappa^{-1/2}C(p,\kappa) = \Theta(e^{-\nu}\nu^{-\frac{1}{2}}) \tag{B.19}$$

converges to 0 exponentially fast.

We now analyze $f_\kappa(s)$ for a given $s \in \mathbb{R}$. Since $s/\sqrt{\kappa} < \pi/2$ for sufficiently large κ, it is sufficient to inspect $f_\kappa(s) = \kappa^{-1/2}C(p,\kappa)(1 + h_1(s/\sqrt{\kappa}))$, where h_1 is the expression in the first line of (3.3) in Theorem 3.2. By (B.19) and $\nu = u - \frac{1}{2}$, we have

$$\begin{aligned} f_\kappa(s) &= \kappa^{-1/2}C(p,\kappa)h_1(s/\sqrt{\kappa}) + \Theta(e^{-\nu}\nu^{-\frac{1}{2}}) \\ &= \frac{1}{\sqrt{2\pi}}\frac{I_u(u\lambda)}{I_{u-\frac{1}{2}}(u\lambda)}\frac{I_u\{u\lambda\cos(s/\sqrt{u\lambda})\}}{I_u(u\lambda)\{\cos(s/\sqrt{u\lambda})\}^{u-1}} + \Theta(e^{-\nu}\nu^{-\frac{1}{2}}). \end{aligned} \tag{B.20}$$

To compare $g_k(s;w)$ with (B.20), we write an alternative limiting form of $g_k(s;w)$. Using Hankel's expansion for the modified Bessel function in (A.8), and the power series representation of $\cos(t)$, we have, as both u and w tend to

infinity,

$$\begin{aligned} g_k(s;w) &= (\lambda u)^{-1/2}\{2\pi I_0(w)\}^{-1} e^{w\cos(s/\sqrt{\lambda u})} \\ &= \frac{1}{\sqrt{\lambda u}2\pi}\frac{\sqrt{2\pi w}}{e^w} e^{w\left(1-\frac{s^2}{2\lambda u}\right)} + \Theta(u^{-1}) \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{w}{\lambda u}} e^{-\frac{w}{2\lambda u}s^2} + \Theta(u^{-1}). \end{aligned}$$

The proof is completed if the following holds:

$$\frac{I_u(u\lambda)}{I_{u-\frac{1}{2}}(u\lambda)} \asymp \sqrt{\frac{w}{\lambda u}} + \Theta(u^{-1}), \tag{B.21}$$

$$\frac{I_u\{u\lambda\cos(s/\sqrt{u\lambda})\}}{I_u(u\lambda)\{\cos(s/\sqrt{u\lambda})\}^{u-1}} \asymp e^{-\frac{w}{2\lambda u}s^2} + \Theta(u^{-1}), \tag{B.22}$$

where $w = \frac{\lambda^2 u}{1+\sqrt{1+\lambda^2}}$. While the uniform expansion of $I_u(u\lambda)$ is the main tool in verifying (B.21) and (B.22), care is needed for (B.21). In particular, we write $I_{u-\frac{1}{2}}(u\lambda) = I_{u-\frac{1}{2}}\{(u-\frac{1}{2})(\frac{u}{u-\frac{1}{2}}\lambda)\}$ to make sure the order $u-\frac{1}{2}$ of the modified Bessel function appears exactly in the argument.

For (B.21), applying the uniform expansion to both Bessel functions gives

$$\begin{aligned} \frac{I_u(u\lambda)}{I_{u-\frac{1}{2}}(u\lambda)} &= \frac{I_u(u\lambda)}{I_{u-\frac{1}{2}}\{(u-\frac{1}{2})(\frac{u}{u-\frac{1}{2}}\lambda)\}} \\ &= \frac{e^{u\eta(\lambda)}}{\sqrt{2\pi u}(1+\lambda^2)^{1/4}} \frac{\sqrt{2\pi(u-\frac{1}{2})}(1+a_u^2\lambda^2)^{1/4}}{e^{(u-\frac{1}{2})\eta(a_u\lambda)}} + \Theta(u^{-1}) \\ &\asymp e^{u\{\eta(\lambda)-\eta(a_u\lambda)\}+\frac{1}{2}\eta(a_u\lambda)}. \end{aligned} \tag{B.23}$$

where $a_u = u/(u-\frac{1}{2}) = 1 + 1/(2u-1)$. A calculation involving Laurant series of the exponent of (B.23) at $u=\infty$ leads that

$$e^{u(\eta(\lambda)-\eta(a_u\lambda))+\frac{1}{2}\eta(a_u\lambda)} \asymp \left(\frac{\lambda}{1+\sqrt{1+\lambda^2}}\right)^{\frac{1}{2}}, \tag{B.24}$$

which shows (B.21). Finally, a similar derivation used for (B.21) can be used to show (B.22). □

For completeness, we provide our proofs for the claims (B.24) and (B.22).

A proof of (B.24), as mentioned in the proof of Theorem B.6. Recall that

$$\eta(\lambda) = \sqrt{1+\lambda^2} + \log\lambda - \log(1+\sqrt{1+\lambda^2}),$$

and that $a_u = 1 + 1/(2u-1)$. Also note that for any constant $a>0$, as $x\to\infty$, $\sqrt{a+\frac{1}{x}+O(x^{-2})} = \sqrt{a} + (2\sqrt{a}x)^{-1} + O(x^{-2})$ and $x\log\{1+\frac{a}{x}+O(x^{-2})\} =$

$a + O(x^{-1})$. Then as $u \to \infty$, we have

$$\sqrt{1 + a_u^2 \lambda^2} = \sqrt{1 + \lambda^2} + \frac{\lambda^2}{\sqrt{1 + \lambda^2}} \frac{1}{2u - 1} + O(u^{-2}),$$
$$u \log(1 + \frac{1}{2u - 1}) = \frac{1}{2} + O(u^{-1}),$$
$$u \log \left(\frac{1 + \sqrt{1 + a_u^2 \lambda^2}}{1 + \sqrt{1 + \lambda^2}} \right) = \frac{\lambda^2}{2\sqrt{1 + \lambda^2}(1 + \sqrt{1 + \lambda^2})} + O(u^{-1}).$$

Combining the above with $\lim_{u \to \infty} \eta(a_u \lambda) = \eta(\lambda)$,

$$\begin{aligned} &\exp \left\{ u(\eta(\lambda) - \eta(a_u \lambda)) + \frac{1}{2} \eta(a_u \lambda) \right\} \\ &\quad \asymp \exp \left\{ -\frac{\lambda^2}{2\sqrt{1 + \lambda^2}} - \frac{1}{2} + \frac{\lambda^2}{2\sqrt{1 + \lambda^2}(1 + \sqrt{1 + \lambda^2})} + \frac{1}{2} \eta(\lambda) \right\} \\ &\quad = \exp \left\{ \frac{1}{2} \log \left(\frac{\lambda}{1 + \sqrt{1 + \lambda^2}} \right) \right\}, \end{aligned}$$

as desired. □

A proof of (B.22), as mentioned in the proof of Theorem B.6. Continuing the proof of Theorem B.6, we treat $s \in \mathbb{R}$ as fixed. Let $r_u(s) = \cos(s/\sqrt{u\lambda})$. Use the uniform expansion of the modified Bessel function as $u \to \infty$ to write

$$\begin{aligned} \frac{I_u\{u\lambda r_u(s)\}}{r_u^{u-1}(s) I_u(u\lambda)} &= \exp\left(u[\eta\{\lambda r_u(s)\} - \eta(\lambda) + \log r_u]\right) r_u(s) \left(\frac{1 + \lambda^2}{1 + \lambda^2 r_u^2(s)} \right)^{\frac{1}{4}} \\ &\quad + \Theta(u^{-1}) \\ &\asymp \exp\left(u[\eta\{\lambda r_u(s)\} - \eta(\lambda) + \log r_u]\right) + \Theta(u^{-1}), \end{aligned}$$

where the second asymptotic equality is given by $\lim_{u \to \infty} r_u(s) = 1$.

We shall use the following large u approximations: As $u \to \infty$,

$$\sqrt{1 + \lambda^2 r_u^2(s)} = \sqrt{1 + \lambda^2} - \frac{\lambda}{2\sqrt{1 + \lambda^2}} \frac{s^2}{u} + O(u^{-2}),$$
$$u \log r_u(s) = -\frac{s^2}{2\lambda} - \frac{s^4}{4! \lambda^2 u} + O(u^{-2}).$$

Then,

$$\begin{aligned} &\exp\left(u[\eta\{\lambda r_u(s)\} - \eta(\lambda) + \log r_u]\right) \\ &\quad = \exp \left\{ -\frac{\lambda}{2\sqrt{1 + \lambda^2}} s^2 + \frac{\lambda}{2\sqrt{1 + \lambda^2}(1 + \sqrt{1 + \lambda^2})} s^2 + O(u^{-1}) \right\} \\ &\quad \asymp \exp \left\{ -\frac{\lambda}{2(1 + \sqrt{1 + \lambda^2})} s^2 \right\}, \end{aligned}$$

as needed. □

B.3. Proofs and technical details for Section 4.2

B.3.1. Low concentration approximation

Proof of Theorem 4.2. We begin with the expression

$$f_{\mathrm{PvM}}(t;p,\kappa,\delta)=\int_0^1 f_R(r)f_{\mathrm{vM}}(t;\kappa\cos(\delta)r)dr$$

given in (3.2). Since $I_\nu^*(\kappa)$ is increasing, $f_R(r)$ is bounded by

$$f_R^{(-)}(r)=\frac{2}{I_\nu^*(\kappa)}I_0^*(0)r(1-r^2)^{\nu-1}I_{\nu-1}^*(0)=\frac{2}{I_\nu^*(\kappa)\Gamma(\nu)}r(1-r^2)^{\nu-1}$$

and

$$f_R^{(+)}(r)=\frac{2}{I_\nu^*(\kappa)}I_0^*\{\kappa\cos(\delta)\}r(1-r^2)^{\nu-1}I_{\nu-1}^*\{\kappa\sin(\delta)\}.$$

That is, uniformly over $r\in[0,1]$, $f_R^{(-)}(r)\le f_R(r)\le f_R^{(+)}(r)$. Furthermore, as $I_0(z)$ is also an increasing function, for any given t, $f_{\mathrm{vM}}(t;\kappa\cos(\delta)r)=[2\pi I_0\{\kappa\cos(\delta)r\}]^{-1}\exp\{\kappa\cos(\delta)\cos(t)r\}$ is bounded by

$$\frac{e^{\kappa\cos(\delta)\cos(t)r}}{2\pi I_0(\kappa\cos\delta)}\le f_{\mathrm{vM}}(t;\kappa\cos(\delta)r)\le\frac{e^{\kappa\cos(\delta)\cos(t)r}}{2\pi I_0(0)},$$

uniformly over $r\in[0,1]$.

Write

$$f^{(+)}(t)=\int_0^1 f_R^{(+)}(r)\frac{e^{\kappa\cos(\delta)\cos(t)r}}{2\pi I_0(0)}dr,$$
$$f^{(-)}(t)=\int_0^1 f_R^{(-)}(r)\frac{e^{\kappa\cos(\delta)\cos(t)r}}{2\pi I_0(\kappa\cos(\delta))}dr.$$

We then have for any $t\in(-\pi,\pi]$,

$$f^{(-)}(t)\le f_{\mathrm{PvM}}(t)\le f^{(+)}(t). \tag{B.25}$$

Rewriting $f^{(+)}(t)$, $f^{(+)}(t)=U(\nu,\kappa,\delta)h_0(t;\nu,\kappa\cos\delta)$, where

$$U(\nu,\kappa,\delta)=\frac{2I_0^*(\kappa\cos\delta)I_{\nu-1}^*(\kappa\sin\delta)}{2\pi I_\nu^*(\kappa)I_0(0)\cdot 2\nu},$$
$$h_0(t;\nu,\kappa\cos\delta)=2\nu\int_0^1 r(1-r^2)^{\nu-1}e^{\kappa\cos(\delta)\cos(t)r}dr.$$

Using the standard power series representation of I_ν (A.1), as $\kappa\to 0$,

$$U(\nu,\kappa,\delta)=1/(2\pi)+O(\kappa^2).$$

Integration by parts gives that $h_0(t;\nu,\kappa\cos\delta) = 1 + 2\nu g(t)$, where

$$g(t) = \frac{\kappa\cos(\delta)\cos(t)}{2\nu}\int_0^1 e^{\kappa\cos(\delta)\cos(t)r}(1-r^2)^\nu dr,$$

which is exactly the expression (B.6) in the proof of Theorem 3.2 with κ replaced by $\kappa\cos(\delta)$. Following the lines of the proof of Theorem 3.2, we have

$$h_0(t;\nu,\kappa\cos\delta) = 1 + h(t;\delta),$$

where $h(t;\delta)$ is the function $h(t)$ in (3.3) in the statement of Theorem 3.2 with κ replaced by $\kappa\cos(\delta)$.

Calculation similar to above leads that $f^{(-)}(t) = L(\nu,\kappa,\delta)h_0(t;\nu,\kappa\cos\delta)$, where

$$L(\nu,\kappa,\delta) = \frac{2}{2\pi I_\nu^*(\kappa)\Gamma(\nu)I_0(\kappa\cos\delta)\cdot 2\nu} = 1/(2\pi) + O(\kappa^2). \tag{B.26}$$

Summarizing the arguments from (B.25) to (B.26), and the fact that $h_0(t;\nu,\kappa\cos\delta)$ is uniformly bounded below and above on $t\in[-\pi,\pi]$,

$$f_{\mathrm{PvM}}(t;\nu,\kappa,\delta) = \frac{1}{2\pi}(1+h(t;\delta)) + O(\kappa^2).$$

The proof is completed by observing

$$1 + h(t;\delta) - \exp(\tilde\kappa\cos t) = \left[\frac{4}{p} - \frac{1}{2}\{B(\tfrac{p}{2},\tfrac{1}{2})\}^2\right]\left(\frac{\kappa\cos\delta\cos t}{2}\right)^2 + O(\kappa^3),$$

for any $t\in(-\pi,\pi]$ as done in (B.17) in the proof of Theorem B.4. □

B.3.2. High dimensional approximations

We give a proof of (4.1), which states that for fixed $\kappa > 0$ and $\delta = \rho(\gamma,\boldsymbol{\mu}) \in [0,\pi/2]$, if $\mathbf{x}\sim \mathrm{vMF}_p(\boldsymbol{\mu},\kappa)$, then $(\sqrt{p}R, S_\gamma(\mathbf{x})) \Rightarrow (X,U)$, as $p\to\infty$, where $X\sim$ Rayleigh(1) and $U\sim$ Uniform$(-\pi,\pi)$ are independent.

Proof of (4.1). Let $\nu = (p-2)/2$ and $Y_\nu = \sqrt{\nu}R$. For $T = S_\gamma(\mathbf{x})$, it is enough to show that $(Y_\nu,T)\Rightarrow(Y,U)$ as $\nu\to\infty$, where $Y\sim$ Rayleigh(1/2) and $U\sim$ Uniform$(-\pi,\pi)$ are independent.

From (B.4) in the proof of Theorem 3.1, (Y_ν,T) has the density function $f(y,t) = \vartheta(\kappa,\nu,d,y)\exp\{\kappa\cos(d)\cos(t)y/\sqrt{\nu}\}\frac{y}{\nu}\left(1-\frac{y^2}{\nu}\right)^{\nu-1}$, where for $z = \kappa\sin(d)\sqrt{1-y^2/\nu}$,

$$\begin{aligned}\vartheta(\kappa,\nu,d,y) &= c_p(\kappa)S_{2\nu-1}I_{\nu-1}(z)\Gamma(\nu)\left(\frac{z}{2}\right)^{-(\nu-1)}\\ &= \frac{1}{\pi}\frac{I_{\nu-1}(z)}{I_\nu(\kappa)}\frac{\kappa}{2}z^{-(\nu-1)}.\end{aligned}$$

By applying (A.9), we have $\vartheta(\kappa, \nu, d, y) \asymp \nu/\pi$ as $\nu \to \infty$. Combining this with $\lim_{\nu\to\infty} \exp\{\kappa \cos(d) \cos(t) y/\sqrt{\nu}\} = 1$ and $\frac{y}{\nu}\left(1 - \frac{y^2}{\nu}\right)^{\nu-1} \asymp \frac{y}{\nu} e^{-s^2}$, we have, for any given $y > 0$, $t \in (-\pi, \pi)$, $\lim_{\nu\to\infty} f(y, t) = 2ye^{-y^2} \cdot \frac{1}{2\pi}$. By Scheffe's Lemma (Durrett, 2019, p. 117), the assertion follows. □

The next two lemmas are used in the proof of Theorem 4.3.

Lemma B.7. *Let $u = p/2 - 2$ for $p \geq 5$. Suppose both u and κ increase while $\lambda = \kappa/u \in (0, \infty)$ is fixed. Suppose that $\mathbf{x} \sim \mathrm{vMF}_p(\boldsymbol{\mu}, \kappa)$, and $\delta = \rho(\boldsymbol{\mu}, \gamma)$ is fixed and $\delta \in [0, \pi/2)$. Then for $(R, S_\gamma(\mathbf{x}))$ as in Theorem 3.1, and*

$$r_0 = \frac{\lambda \cos\delta}{1 + \sqrt{1 + \lambda^2}},$$

$\sqrt{p}(R - r_0)$ is asymptotically normally distributed with mean 0, as $u \to \infty$.

Proof of Lemma B.7. Write $\kappa = u\lambda$ and

$$\omega(r) = \lambda \sin(\delta)\sqrt{1 - r^2}.$$

By Theorem 3.2,

$$f_R(r) = c_1 I_0\{u\lambda \cos(\delta) r\} r(1 - r^2)^u \frac{I_u\{u\omega(r)\}}{\{\omega(r)\}^u}.$$

Throughout, $c_1, c_2, c_3, \ldots$ denote the appropriate normalizing constants.

Using the Hankel's expansion for I_0 in (A.8), we write

$$I_0\{u\lambda \cos(d) r\} r(1 - r^2)^u = \exp\left[u\{\lambda \cos(d) r + \log(1 - r^2)\}\right] M_1(r),$$

where $M_1(r) = r\{2\pi\lambda \cos(d) r\}^{-1/2}(1 + O(u^{-1}))$. Likewise, using the uniform expansion of the modified Bessel function in (A.13),

$$\frac{I_u\{u\omega(r)\}}{\{\omega(r)\}^u} = \exp\left[u\eta\{\omega(r)\} - u \log\{\omega(r)\}\right] M_2(r),$$

where η is defined in (A.14) and $M_2(r) = \{2\pi\omega(r)\}^{-1/2}\{1 - \omega^2(r)\}^{-1/4}\{1 + O(u^{-1})\}$. Combining the above we have

$$f_R(r) = c_2 \exp\{-u\beta(r)\} M(r),$$

where $M(r) = M_1(r)M_2(r)$ and

$$\beta(r) = -\lambda \cos(d) r - \sqrt{1 + \omega^2(r)} + \log\{1 + \sqrt{1 + \omega^2(r)}\} - \log(1 - r^2). \quad \text{(B.27)}$$

In Lemma B.8, stated below, we confirm that β is a convex function with a unique minimum at $r_0 = \cos(d)(\sqrt{\lambda^2 + 1} - 1)/\lambda \in [0, 1)$.

Let $Y_u = \sqrt{u}(R - r_0)$. The density function of Y_u is $f_u(y) = c_3 \exp\{-u\beta(r_0 + y/\sqrt{u})\} M(r_0 + y/\sqrt{u})$. Since $\delta < \pi/2$, $\cos\delta > 0$, thus $r_0 > 0$. Using this

fact, it can be shown that $\lim_{u\to\infty} M(r_0 + y/\sqrt{u})$ is a non-zero constant. Since $\beta^{(1)}(r_0) = 0$, a Taylor series expansion of $\beta(r)$ at $r = r_0$ gives, as $u \to \infty$,

$$f_u(y) \asymp c_4 \exp\{-\beta^{(2)}(r_0)\frac{y^2}{2}\}.$$

Thus, $f_u(y)$ converges pointwise to $\phi(y/\sigma_{r_0})/\sigma_{r_0}$, $\sigma_{r_0}^2 = \{\beta^{(2)}(r_0)\}^{-1}$, as $u \to \infty$, and we have

$$Y_u \Rightarrow N(0, \sigma_{r_0}^2). \qquad \square$$

Lemma B.8. *For $\delta \in [0, \pi/2)$ and $\lambda > 0$, let $\beta : [0,1] \to \mathbb{R}$ be the function defined in (B.27). Then, β is convex and has a unique minimum at $r_0 = \cos\delta(\sqrt{\lambda^2+1} - 1)/\lambda \in (0,1)$. Moreover, the second derivative of β is strictly positive for all $r \in (0,1)$.*

Proof of Lemma B.8. Calculations show that the first two derivatives of β are

$$\beta^{(1)}(r) = -\lambda\cos\delta + \frac{r}{1-r^2}\{1 + \sqrt{1+\lambda^2(\sin\delta)^2(1-r^2)}\},$$
$$\beta^{(2)}(r) = \frac{\{1 + \sqrt{1+\lambda^2(\sin\delta)^2(1-r^2)}\}(1+r^2) + \lambda^2(\sin\delta)^2(1-r^2)}{(1-r^2)^2\sqrt{1+\lambda^2(\sin\delta)^2(1-r^2)}}.$$

It is easy to confirm that $\beta^{(2)}(r) > 0$ for all $r \in [0,1)$, and thus β is convex. Moreover, the algebraic equalities

$$\begin{aligned} 1 + \lambda^2(\sin\delta)^2(1-r_0^2) &= \{(\sin\delta)^2\sqrt{1+\lambda^2} + (\cos\delta)^2\}^2, \\ 1 + \sqrt{1+\lambda^2(\sin\delta)^2(1-r_0^2)} &= 2 + (\sin\delta)^2(\sqrt{1+\lambda^2}+1), \\ \frac{r_0}{1-r_0^2} &= \frac{\lambda\cos\delta}{2 + (\sin\delta)^2(\sqrt{1+\lambda^2}+1)} \end{aligned}$$

lead that $\beta^{(1)}(r_0) = 0$. Since β is convex, r_0 is the unique minimum of β. Finally, since $(\sqrt{\lambda^2+1} - 1)/\lambda \in (0,1)$ for all $\lambda > 0$, we confirm that $r_0 \in (0,1)$. $\square$

Proof of Theorem 4.3. Let $T = S_\gamma(\mathbf{x})$ and R be as given in Theorem 3.1 so that given $R = r$, $T \sim \mathrm{vM}(\kappa\cos(\delta)r)$.

Suppose $\cos(\delta) = 0$. Then for any p, κ, $T \sim \mathrm{vM}(0)$, which is Uniform$(-\pi, \pi)$.

Assume $\cos(\delta) > 0$ hereafter. By Lemma B.2, we have for any $r \in (0,1)$,

$$\sqrt{\kappa\cos(\delta)r}T \mid (R = r) \Rightarrow N(0,1), \tag{B.28}$$

as $\kappa \to \infty$.

Note that $\tilde{\kappa} = u\lambda\cos(\delta)r_0$. Let $S = \sqrt{\tilde{\kappa}}T = \sqrt{u\lambda\cos(\delta)r_0}T$. Then, for any $x \in \mathbb{R}$,

$$\begin{aligned} \Pr(S \le x) &= \mathrm{E}\{\Pr(\sqrt{u\lambda\cos(\delta)r_0}T \le x \mid R)\} \\ &= \mathrm{E}[\Pr\{(u\lambda\cos(\delta)R)^{1/2}T \le x\sqrt{R/r_0} \mid R\}]. \end{aligned}$$

For the given x, let $X_u(R) = \Pr\{(u\lambda\cos(d)R)^{1/2}T \le x\sqrt{R/r_0} \mid R\}$. Due to (B.28), $\lim_{u\to\infty} X_u(r) = \Phi(x\sqrt{r/r_0})$ for any $r \in (0,1)$. As shown in Lemma B.7, $R \to r_0$ in probability as $u \to \infty$, and we have that $X_u(R) \to \Phi(x\sqrt{r_0/r_0}) = \Phi(x)$ in probability.

Since the sequence of random variables $\{X_u(R)\}$ (indexed by u) is uniformly integrable (as $0 \le X_u(R) \le 1$), and $X_u(R) \to \Phi(x)$ in probability, we have $\mathrm{E}|X_u(R)| - \mathrm{E}|\Phi(x)| \to 0$ as $u \to \infty$, by, e.g., Theorem 5.5.2 of Durrett (2019). Since both terms are nonnegative, we have, for any $x \in \mathbb{R}$, $\Pr(S \le x) \to \Phi(x)$, that is, $S = \sqrt{\tilde{\kappa}}T \Rightarrow N(0,1)$ as $u \to \infty$. □

Proof of Theorem 4.4. Let $(X_{1,p}, X_{2,p}) = (X_1, X_2) = (\mathbf{q}'\mathbf{x}_p, \mathbf{v}'\mathbf{x}_p)$. In this proof the subscript p is sometimes omitted if the context is clear. Using the intermediate result (B.3) in the proof of Theorem 3.1, the conditional density of (X_1, X_2) given $(\mathbf{q}, \mathbf{v})$ is

$$f(x_1, x_2 \mid (\mathbf{q}, \mathbf{v})) = c_p(\sqrt{p}\lambda)S_{p-3}e^{\sqrt{p}\lambda\cos\Delta_p x_1}(1 - x_1^2 - x_2^2)^{\nu-1}I^*_{\nu-1}(x)\Gamma(\nu),$$

for $x_1^2 + x_2^2 \le 1$, where $\nu = (p-2)/2$, $x = \sqrt{p}\lambda\sin\Delta_p\sqrt{1 - x_1^2 - x_2^2}$ and $I^*_\nu(z) = (\frac{z}{2})^{-\nu}I_\nu(z)$. Notice that the dependence on (X_1, X_2) of $(\mathbf{q}, \mathbf{v})$ is only through $\Delta_p = \cos^{-1}(\mathbf{q}'\boldsymbol{\mu}_p)$. Thus, $f(x_1, x_2 \mid (\mathbf{q}, \mathbf{v})) = f(x_1, x_2 \mid \Delta_p)$.

Fix any $\delta_0 \in [0, \pi/2]$. The conditional density of $(Z_1, Z_2) := \sqrt{p}(X_1, X_2)$ given $\Delta_p = \delta_0$ is then

$$f(z_1, z_2 \mid \delta_0) = \zeta_1(z_1, z_2; p, \lambda, \delta_0) \cdot \zeta_2(z_1, z_2; p, \lambda, \delta_0),$$

where

$$\zeta_1(z_1, z_2; p, \lambda, \delta_0) = e^{\lambda\cos\delta_0 z_1}\left(1 - \tfrac{z_1^2+z_2^2}{p}\right)^{\frac{p}{2}-2},$$

and

$$\zeta_2(z_1, z_2; p, \lambda, \delta_0) = c_p(\sqrt{p}\lambda)S_{p-3}I^*_{\nu-1}\left(\lambda\sin\delta_0\sqrt{1 - \tfrac{z_1^2+z_2^2}{p}}\right)\Gamma(\nu)p^{-1}.$$

Then for any $(z_1, z_2) \in \mathbb{R}^2$,

$$\begin{aligned}\lim_{p\to\infty}\zeta_1(z_1, z_2; p, \lambda, \delta_0) &= e^{\lambda\cos\delta_0 z_1}e^{-\frac{1}{2}(z_1^2+z_2^2)} \\ &= e^{-\frac{1}{2}\{(z_1-\lambda\cos\delta_0)^2+z_2^2\}}e^{\frac{1}{2}(\lambda^2\cos^2\delta_0)},\end{aligned}$$

and $\lim_{p\to\infty}\zeta_2(z_1, z_2; p, \lambda, \delta_0)e^{\frac{1}{2}(\lambda^2\cos^2\delta_0)} = (2\pi)^{-1}$ can be shown using the uniform expansion of the modified Bessel function (A.13) and the Poincaré expansion of the gamma function (A.12). By Scheffe's Lemma, $\lim_{p\to\infty} f(z_1, z_2 \mid \delta_0) = \frac{1}{2\pi}e^{-\frac{1}{2}\{(z_1-\lambda\cos\delta_0)^2+z_2^2\}}$ for any z_1, z_2 leads that as $p \to \infty$,

$$\mathcal{L}(Z_1, Z_2 \mid \Delta_p = \delta_0) \Rightarrow N_2((\lambda\cos\delta_0, 0)', \mathbf{I}_2), \tag{B.29}$$

where $\mathcal{L}(X \mid Y = y)$ means the law of X conditioned on $Y = y$.

For each p, write the joint distribution function of (Z_1, Z_2) by $F_p(z_1, z_2)$. Recall that $\Delta_p \to \delta$ in probability as $p \to \infty$ and that δ is deterministic. Then for any z_1, z_2,

$$\begin{aligned} F_p(z_1, z_2) &= \mathrm{E}\{\Pr(Z_1 \le z_2, Z_2 \le z_2 \mid \Delta_p)\} \\ &= \mathrm{E}\{Y_p(\Delta_p)\}, \end{aligned}$$

where $Y_p(\Delta_p) = F_p(z_1, z_2 \mid \Delta_p)$.

Let $\Phi(z_1, z_2 \mid \delta_0)$ be the distribution function of $N_2((\lambda \cos\delta_0, 0)', \mathbf{I}_2)$. By (B.29), $\lim_{p\to\infty} Y_p(\delta_0) = \Phi(z_1, z_2 \mid \delta_0)$. Since $\Delta_p \to \delta$ in probability, we have

$$Y_p(\Delta_p) \to \Phi(z_1, z_2 \mid \delta) \ \text{ in probability.}$$

The sequence $Y_p(\Delta_p)$ is uniformly bounded by 1, thus uniformly integrable. Since the convergence in probability of a uniformly integrable sequence implies the convergence in L_1 (Theorem 5.5.2, Durrett, 2019), we have $F_p(z_1, z_2) = \mathrm{E}\{Y_p(\Delta_p)\} \to \Phi(z_1, z_2 \mid \delta)$ as $p \to \infty$, i.e.,

$$(Z_1, Z_2) \Rightarrow N_2((\lambda \cos\delta, 0)', \mathbf{I}_2), \tag{B.30}$$

as $p \to \infty$.

Recall that $T = S_\gamma(\mathbf{x}_p)$ is the polar angle of (Z_1, Z_2). The mapping $(Z_1, Z_2) \mapsto T$ is continuous except at $(Z_1, Z_2) = \mathbf{0}$, which is a measure zero set with respect to the limiting distribution of (B.30). Then, the continuous mapping theorem gives $S_\gamma(\mathbf{x}_p) \Rightarrow \mathrm{PN}_2((\lambda \cos\delta, 0)', \mathbf{I}_2)$ as $p \to \infty$. □

Proof of Lemma 4.5. (i) The high concentration case. Note that

$$f_\lambda(s) = \frac{1}{2\pi}\lambda^{-1} e^{-\frac{\lambda^2}{2}} \left\{ 1 + \frac{\Phi\{\lambda\cos(\frac{s}{\lambda})\}}{\phi\{\lambda\cos(\frac{s}{\lambda})\}} \lambda \cos(\tfrac{s}{\lambda}) \right\}.$$

Since for any $\alpha > 0$, $\lambda^\alpha e^{-\lambda^2/2} \to 0$ as $\lambda \to \infty$, we have

$$\lambda^2\{f_\lambda(s) - \phi(s)\} = o(1) + \lambda^2 \left[\frac{1}{2\pi} e^{-\frac{\lambda^2}{2}} \frac{\Phi\{\lambda\cos(\frac{s}{\lambda})\}}{\phi\{\lambda\cos(\frac{s}{\lambda})\}} \cos(\tfrac{s}{\lambda}) - \phi(s) \right]. \tag{B.31}$$

By replacing $\cos(s/\lambda)$ with $1 - \frac{s^2}{2\lambda^2} + \frac{s^4}{4!\lambda^4} + O(\lambda^{-6})$, the second term of (B.31) becomes

$$\frac{\lambda^2}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \left[e^{\frac{s^4}{3\lambda^2} + O(\lambda^{-4})} \Phi\{\lambda + O(\lambda^{-1})\} \left\{ 1 - \frac{s^2}{2\lambda^2} + O(\lambda^{-4}) \right\} - 1 \right]. \tag{B.32}$$

By using the normal tail bound, $\Phi(\lambda + O(\lambda^{-1})) = 1 - O(\lambda^{-1} e^{-\lambda^2/2})$ (as $\lambda \to \infty$) and the power series representation of the exponential, (B.32) is

$$\begin{aligned} &\frac{\lambda^2}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \left[\left\{ 1 + \frac{s^4}{3\lambda^2} - \frac{s^2}{2\lambda^2} + O(\lambda^{-4}) \right\} - 1 \right] \\ &\quad = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \left\{ \frac{s^4}{3} - \frac{s^2}{2} + O(\lambda^{-2}) \right\}, \end{aligned}$$

which, when plugged in to (B.31), shows $f_\lambda(s) - \phi(s) = \Theta(\lambda^{-2})$. The proof is completed by noting that Lemma B.2 gives $g_\lambda(s;\lambda^2) = \phi(s) + \Theta(\lambda^{-2})$.

(ii) The low concentration case. Let $\eta(t;\lambda) = \frac{\Phi(\lambda\cos t)}{\phi(\lambda\cos t)}\lambda\cos t$. As $\lambda \to 0$, for $\cos t > 0$,

$$\begin{aligned}\eta(t;\lambda) &= \frac{\frac{1}{2} + \int_0^{\lambda\cos t} 1 - \frac{x^2}{2} + O(x^4)dx}{\frac{1}{\sqrt{2\pi}}e^{-\frac{\lambda^2\cos^2 t}{2}}}\lambda\cos t \\ &= \sqrt{\frac{\pi}{2}}\lambda\cos t + \sqrt{2\pi}\lambda^2\cos^2 t + O(\lambda^3). \end{aligned} \tag{B.33}$$

It can be checked that (B.33) holds for $\cos t \le 0$ as well. Let $\tilde\lambda = \sqrt{\frac{\pi}{2}}\lambda$. Then

$$\begin{aligned}\lambda^{-2}\{f_{\mathrm{PN}}(t;\lambda) - f_{\mathrm{Cardioid}}(t;\tilde\lambda/2)\} &= \frac{\lambda^{-2}}{2\pi}\left[e^{-\frac{\lambda^2}{2}}\{1+h(t;\lambda)\} - (1+\tilde\lambda\cos t)\right] \\ &= \frac{\lambda^{-2}}{2\pi}\left[(e^{-\frac{\lambda^2}{2}}-1)\{1+h(t;\lambda)\} + (h(t;\lambda) - \tilde\lambda\cos t)\right] \\ &= \frac{\lambda^{-2}}{2\pi}\left\{-\frac{\lambda^2}{2}(1+\tilde\lambda\cos t) + O(\lambda^4) + \sqrt{2\pi}\lambda^2\cos^2 t + O(\lambda^3)\right\}.\end{aligned}$$

Thus, $f_{\mathrm{PN}}(t;\lambda) = f_{\mathrm{Cardioid}}(t;\tilde\lambda/2) + \Theta(\lambda^2)$. Applying Lemma B.3, we get

$$f_{\mathrm{Cardioid}}(t;\tilde\lambda/2) = f_{\mathrm{vM}}(t;\tilde\lambda) + \Theta(\lambda^2)$$

and $f_{\mathrm{PN}}(t;\lambda) = f_{\mathrm{vM}}(t;\tilde\lambda) + \Theta(\lambda^2)$ as required. □

B.4. Proof of Corollary 5.1

Proof of Corollary 5.1. Introduce a latent variable W, with $\Pr(W=1) = \pi_1 = 1 - \Pr(W=2)$, such that $\mathbf{x}_p$ given $W = w$ follows $\mathrm{vMF}_p(\boldsymbol{\mu}_w, \kappa_p)$ for $w = 1, 2$. The location parameters $\boldsymbol{\mu}_w = \boldsymbol{\mu}_w(p)$ are unit vectors in $\mathbb{R}^p$. The sequences $\{\boldsymbol{\mu}_w(p) : p = 1, 2, \dots, \}$ for $w = 1, 2$ are deterministic.

(i) For $i = 1, 2$, let $\mathbf{y}_i \sim$ i.i.d. $N_p(\mathbf{0}, I_p)$. Write

$$\mathbf{P} = [\mathbf{q}, \mathbf{v}] = [\mathbf{y}_1, \mathbf{y}_2]\begin{pmatrix}\mathbf{y}_1'\mathbf{y}_1 & \mathbf{y}_1'\mathbf{y}_2 \\ \mathbf{y}_2'\mathbf{y}_1 & \mathbf{y}_2'\mathbf{y}_2\end{pmatrix}^{-\frac{1}{2}}.$$

Then, $\mathbf{P}$ follows the uniform distribution on $V_{2,p}$ (Chikuse, 2012). For the geodesic $\gamma = \gamma(\mathbf{q}, \mathbf{v})$, $\Delta_p(w) = \rho(\boldsymbol{\mu}_w, \gamma) = \cos^{-1}(\|\mathbf{P}'\boldsymbol{\mu}_w\|)$. Note that $\mathbf{q}$ is not in general the closest point to $\boldsymbol{\mu}_w$ on γ. Since

$$\|\mathbf{P}'\boldsymbol{\mu}_w\| = \left\|\left\{\frac{1}{p}\begin{pmatrix}\mathbf{y}_1'\mathbf{y}_1 & \mathbf{y}_1'\mathbf{y}_2 \\ \mathbf{y}_2'\mathbf{y}_1 & \mathbf{y}_2'\mathbf{y}_2\end{pmatrix}\right\}^{-\frac{1}{2}}\begin{pmatrix}\mathbf{y}_1'\boldsymbol{\mu}_w/\sqrt{p} \\ \mathbf{y}_2'\boldsymbol{\mu}_w/\sqrt{p}\end{pmatrix}\right\| \to 0,$$

in probability as $p \to \infty$, $\Delta_p(w) \to \pi/2$ in probability for $w = 1, 2$. Then by Theorem 4.4, conditional to $W = w$, $S_\gamma(\mathbf{x}_p) \Rightarrow \mathrm{PN}_2(\mathbf{0}, \mathbf{I}_2)$ as $p \to \infty$, where

$\mathrm{PN}_2(\mathbf{0}, \mathbf{I}_2)$ is the uniform distribution on $(-\pi, \pi]$. Since the limiting conditional distributions are equal for both $w = 1, 2$, the assertion follows.

(ii) Assume without losing generality, $\boldsymbol{\mu} = \mathbf{e}_1$ and $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}\}$ are contained in the span of $\mathbf{e}_1, \mathbf{e}_2$. Let $d_p = \rho(\boldsymbol{\mu}, \boldsymbol{\mu}_1) = \rho(\boldsymbol{\mu}, \boldsymbol{\mu}_2) \in [0, \pi/2]$. Again without losing generality, we write $\boldsymbol{\mu}_1 = \cos(d_p)\mathbf{e}_1 + \sin(d_p)\mathbf{e}_2$ and $\boldsymbol{\mu}_2 = \cos(d_p)\mathbf{e}_1 - \sin(d_p)\mathbf{e}_2$. Note that $\lim_{p\to\infty} d_p = d$.

With the convention $\boldsymbol{\mu} = \mathbf{e}_1$, the random vector $\mathbf{v}$ is represented by

$$\mathbf{v} = \mathbf{y}/\sqrt{\mathbf{y}'\mathbf{y}}, \ \mathbf{y} = (0, Y_1, \ldots, Y_{p-1})',$$

where $Y_i \sim$ i.i.d. $N(0, 1)$. Then for $w = 1, 2$,

$$\mathbf{P}'\boldsymbol{\mu}_w = \begin{pmatrix} \boldsymbol{\mu}'\boldsymbol{\mu}_w \\ \mathbf{y}'\boldsymbol{\mu}_w/\sqrt{\mathbf{y}'\mathbf{y}} \end{pmatrix} \to \begin{pmatrix} \cos d \\ 0 \end{pmatrix}$$

and $\Delta_p(w) = \cos^{-1}(\|\mathbf{P}'\boldsymbol{\mu}_w\|) \to d$ in probability as $p \to \infty$, Theorem 4.4 leads that conditional to $W = w$, $S_\gamma(\mathbf{x}_p) \Rightarrow \mathrm{PN}_2((\lambda \cos d, 0)', \mathbf{I}_2)$ in probability as $p \to \infty$ for both $w = 1, 2$. Since the limiting conditional distributions are $\mathrm{PN}_2((\lambda \cos d, 0)', \mathbf{I}_2)$ for both $W = 1, 2$, the assertion follows. □

Appendix C: An extended comparison of approximated vM distributions

As referenced in Section 4.3, the qualities of the vM approximations in Section 4 are further inspected in this appendix.

For each triple (p, κ, δ), there are five approximations summarized in Table 1. For notational simplicity, we denote $f_{(i)}$ for the approximated density under the ith scenario of Table 1. The same order is used in the legends of Figs. 6 and C.1. For example, $f_{(1)}$ refers to the "large κ" approximation, shown in the first row of Table 1.

The quality of approximation $f_{(i)}$, measured by the Hellinger distance to $f := f_{\mathrm{PvM}}(p, \kappa, \delta)$, is inspected separately over two regions of interest. In Fig. C.1, for each choice of $\delta \in \{0°, 30°, 60°, 80°\}$, regions "A" and "B" of (κ, p) represent regions in which the $f_{(5)}$ (and $f_{(4)}$, respectively) approximations appear to work best. These regions are numerically obtained, but roughly, region A corresponds to the cases $\kappa < 2p$, and region B to $\kappa > 2p$. For region A, the projected normal approximation $f_{(5)}$ is chosen as a reference, and we evaluate the difference of the quality by

$$D(i\|\text{reference}) := H(f, f_{(\text{reference})}) - H(f, f_{(i)}),$$

over a fine grid of $(\kappa, p) \in A$ and for all four choices of δ. If $D(i\|\text{reference}) > 0$, then the approximation of $f_{(i)}$ is better than that of $f_{(\text{reference})}$ (and vice versa). The higher $D(i\|\text{reference})$, the better approximation of $f_{(i)}$ over $f_{(\text{reference})}$. In

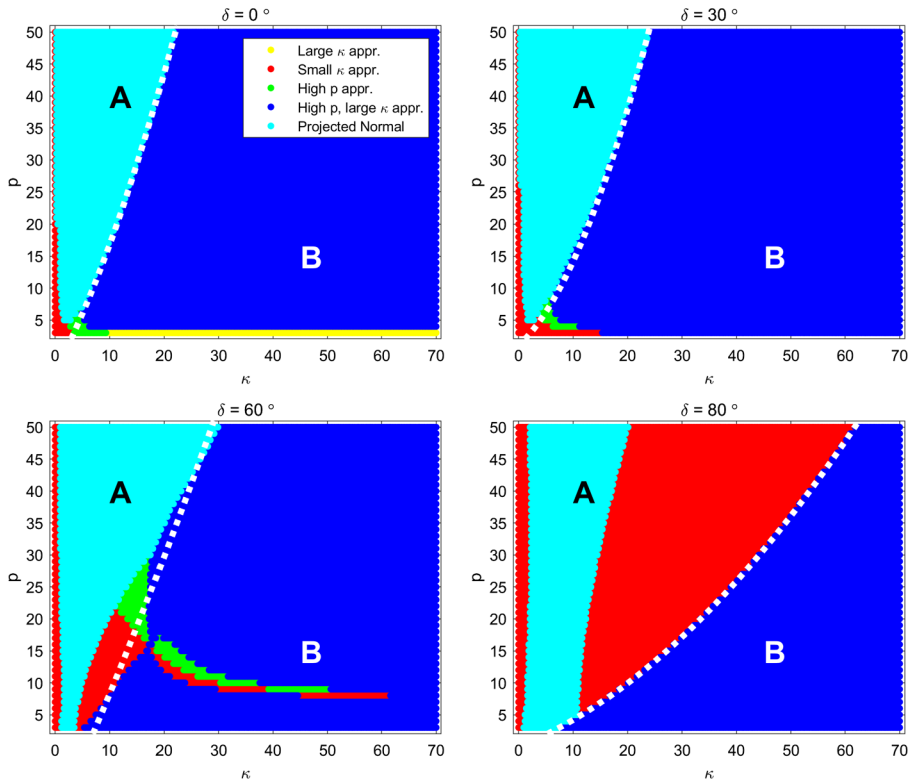

FIG C.1. *Extension of Fig. 6. Regions A and B of* (κ, p) *are split by the white dashed curve.*

Fig. C.2, left column, the histograms of $D(i\|\text{reference})$ values over region A are shown for the five approximations ($i = 1, \ldots, 5$). The right column of the figure is similarly obtained for region B, with $f_{(4)}$ (high p, high κ) as the reference. We observe the following:

- For region A ($\kappa < 2p$), approximations given by $f_{(2)}$, $f_{(3)}$ and $f_{(5)}$ are of similar quality. (Differences in Hellinger distance are at most 0.059.)
- For region B ($\kappa > 2p$), the approximation by $f_{(4)}$ (high p, high κ) shows the best quality. The projected normal approximation by $f_{(5)}$ is among the worst.
- Overall, $f_{(4)}$ provides the best approximation.

The histograms in Figure C.2 are obtained for all choices of δ. For completeness, Table C.1 records the minimum, median and maximum of $D(i\|\text{reference})$ for each choice of δ and for all choices combined. The patterns are similar across different values of δ.

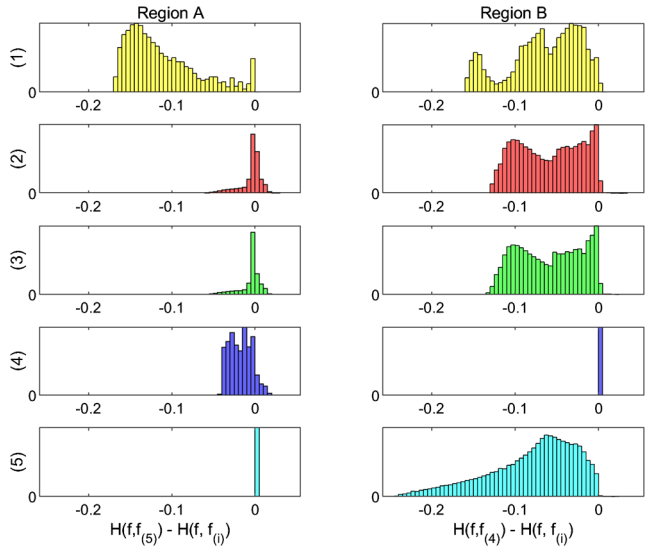

FIG C.2. *The quality of approximations compared to the reference. Shown are the histograms of* $D(i\|\text{reference})$. *The* i*th row of the figure corresponds to the* i*th approximation. For region A,* $f_{(2)}$, $f_{(3)}$ *and* $f_{(5)}$ *are of similar quality. For region B, the quality of* $f_{(4)}$ *approximation stands out.*

TABLE C.1
(Minimum, median, maximum) of $100 \times D(i\|\text{reference})$ *over* $(\kappa, p) \in A$ *or* B.

δ	Scenario	Region A			Region B		
0°	(1)	(-16.8 ,	-10.8 ,	0)	(-8.3 ,	-2.7 ,	0.6)
	(2)	(-5.9 ,	-1.4 ,	1.1)	(-12.9 ,	-9.4 ,	1.8)
	(3)	(-5.9 ,	-1.4 ,	1.0)	(-12.9 ,	-9.3 ,	1.8)
	(4)	(-4.3 ,	-1.5 ,	0.4)	(0 ,	0 ,	0)
	(5)	(0 ,	0 ,	0)	(-23.7 ,	-7.2 ,	0)
30°	(1)	(-16.8 ,	-11.1 ,	0)	(-8.8 ,	-4.0 ,	0.8)
	(2)	(-4.9 ,	-1.2 ,	0.3)	(-11.7 ,	-8.0 ,	3.3)
	(3)	(-4.9 ,	-1.1 ,	0.3)	(-11.6 ,	-7.9 ,	2.1)
	(4)	(-4.4 ,	-1.4 ,	0.5)	(0 ,	0 ,	0)
	(5)	(0 ,	0 ,	0)	(-24.2 ,	-7.3 ,	2.3)
60°	(1)	(-16.8 ,	-12.5 ,	0.1)	(-11.9 ,	-8.4 ,	-5.0)
	(2)	(-1.0 ,	-0.2 ,	2.7)	(-5.7 ,	-2.0 ,	0.3)
	(3)	(-1.4 ,	-0.2 ,	1.9)	(-5.6 ,	-2.1 ,	0.2)
	(4)	(-4.1 ,	-2.1 ,	4.6)	(0 ,	0 ,	0)
	(5)	(0 ,	0 ,	0)	(-23.8 ,	-6.6 ,	0.1)
80°	(1)	(-16.7 ,	-13.2 ,	0)	(-15.9 ,	-14.3 ,	-2.6)
	(2)	(-0.4 ,	0.1 ,	1.7)	(-12.9 ,	-2.8 ,	0)
	(3)	(-1.4 ,	0 ,	1.7)	(-13.4 ,	-3.0 ,	-0.1)
	(4)	(-3.3 ,	-1.9 ,	1.7)	(0 ,	0 ,	0)
	(5)	(0 ,	0 ,	0)	(-22.4 ,	-5.9 ,	0)
overall	(1)	(-16.8 ,	-12.3 ,	0.1)	(-15.9 ,	-5.9 ,	0.8)
	(2)	(-5.9 ,	-0.0 ,	2.7)	(-12.9 ,	-5.6 ,	3.3)
	(3)	(-5.8 ,	-0.1 ,	1.9)	(-13.4 ,	-5.5 ,	2.1)
	(4)	(-4.4 ,	-1.7 ,	4.6)	(0 ,	0 ,	0)
	(5)	(0 ,	0 ,	0)	(-24.2 ,	-6.9 ,	2.3)

References

Alashwali, F. and Kent, J. T. (2016). The use of a common location measure in the invariant coordinate selection and projection pursuit. *Journal of Multivariate Analysis* **152** 145–161. MR3554783

Bickel, P. J., Kur, G. and Nadler, B. (2018). Projection pursuit in high dimensions. *Proceedings of the National Academy of Sciences* **115** 9151–9156. MR3856112

Chikuse, Y. (2012). *Statistics on special manifolds* **174**. Springer Science & Business Media. MR1960435

Cornea, E., Zhu, H., Kim, P. and Ibrahim, J. G. (2017). Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 463–482. MR3611755

Damon, J. and Marron, J. S. (2014). Backwards principal component analysis and principal nested relations. *Journal of Mathematical Imaging and Vision* **50** 107–114. MR3233137

Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *The Annals of Statistics* **12** 793–815. MR0751274

Dingle, R. B. (1973). *Asymptotic Expansions: Their Derivation and Interpretation*. Academic Press London. MR0499926

DLMF (2020). *NIST Digital Library of Mathematical Functions*. `http://dlmf.nist.gov/`, Release 1.0.28 of 2020-09-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *The Annals of Statistics* **33** 1643–1665. MR2166558

Durrett, R. (2019). *Probability: Theory and Examples* **49**. Cambridge university press. MR3930614

Fletcher, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision* **105** 171–185. MR3104017

Fletcher, P. T., Lu, C., Pizer, S. M. and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging* **23** 995–1005.

Friedman, J. H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association* **82** 249–266. MR0883353

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association* **76** 817–823. MR0650892

Friedman, J. H. and Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on computers* **100** 881–890.

Huber, P. J. (1985). Projection pursuit. *The Annals of Statistics* **13** 435–475. MR0790553

Huckemann, S. F. and Eltzner, B. (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *The Annals of Statistics* **46** 1994–2019. MR3845008

Jee, J. R. (2009). Projection pursuit. *Wiley Interdisciplinary Reviews: Computational Statistics* **1** 208–215.

Jones, M. and Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association* **100** 1422–1428. MR2236452

Jung, S., Dryden, I. L. and Marron, J. S. (2012). Analysis of Principal Nested Spheres. *Biometrika* **99** 551–568. MR2966769

Kent, J. T. (1978). Limiting behaviour of the von Mises-Fisher distribution. In *Mathematical Proceedings of the Cambridge Philosophical Society* **84** 531–536. Cambridge University Press. MR0501259

Kent, J. T. (1982). The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society. Series B (Methodological)* **44** 71–80. MR0655376

Kim, B., Schulz, J. and Jung, S. (2020). Kurtosis test of modality for rotationally symmetric distributions on hyperspheres. *Journal of Multivariate Analysis* **178** 104603. MR4076342

Kim, B., Huckemann, S., Schulz, J. and Jung, S. (2019). Small-sphere distributions for directional data with application to medical imaging. *Scandinavian Journal of Statistics* **46** 1047–1071. MR4033803

Kuiper, N. H. (1960). Tests concerning random points on a circle. In *Nederl. Akad. Wetensch. Proc. Ser. A* **63** 38–47. MR0111088

Ley, C. and Verdebout, T. (2017). *Modern directional statistics*. CRC Press. MR3752655

Mardia, K. V. (1975). Distribution theory for the von Mises-Fisher distribution and its application. In *A Modern Course on Statistical Distributions in Scientific Work* 113–130. Springer. MR0415862

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley. MR1828667

Nemes, G. (2017). Error bounds for the large-argument asymptotic expansions of the Hankel and Bessel functions. *Acta Applicandae Mathematicae* **150** 141–177. MR3668093

Olver, F. W. J. (1997). *Asymptotics and special functions*. AK Peters/CRC Press. MR1429619

Olver, F. W. J., Lozier, D. W., Boisvert, R. F. and Clark, C. W. (2010). *NIST handbook of mathematical functions hardback and CD-ROM*. Cambridge university press. MR2655349

Paindaveine, D. and Verdebout, T. (2020). Inference for spherical location under high concentration. *The Annals of Statistics* **48** 2982–2998. MR4152631

Pennec, X. et al. (2018). Barycentric subspace analysis on manifolds. *The Annals of Statistics* **46** 2711–2746. MR3851753

Pewsey, A., Neuhäuser, M. and Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press. MR3156170

Pizer, S. M. and Marron, J. S. (2017). Object statistics on curved manifolds. In *Statistical Shape and Deformation Analysis* 137–164. Elsevier.

Presnell, B., Morrison, S. P. and Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association* **93** 1068–1077. MR1649201

Pukkila, T. M. and Rao, C. R. (1988). Pattern recognition based on scale invariant discriminant functions. *Information sciences* **45** 379–389. MR0952440

Qi, F. (2010). Bounds for the ratio of two gamma functions. *Journal of Inequalities and Applications* **2010** 1–84. MR2611044

Scealy, J. L. and Wood, A. T. A. (2019). Scaled von Mises–Fisher Distributions and Regression Models for Paleomagnetic Directional Data. *Journal of the American Statistical Association* **114** 1547–1560. MR4047280

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. MR3357394

Shimizu, K. and Iida, K. (2002). Pearson type VII distributions on spheres. *Communications in Statistics-Theory and Methods* **31** 513–526. MR1902308

Watson, G. S. (1983). *Statistics on spheres. University of Arkansas lecture notes in the mathematical sciences*. Wiley, New York. MR0709262

Watson, G. S. (1988). The Langevin distribution on high dimensional spheres. *Journal of Applied Statistics* **15** 123–130.