

On parameter estimation for $N(\mu, \sigma^2 I_3)$ based on projected data into \mathbb{S}^2

Jordi-Lluís Figueras^a, Aron Persson^{a,*}, Lauri Viitasaari^b

^a*Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden*

^b*Department of Information and Service Management, Aalto University School of Business, PO Box 11110, 00076 Aalto, Finland*

jordi-lluis.figueras@math.uu.se (J. Figueras), aron.persson@math.uu.se (A. Persson),
lauri.viitasaari@aalto.fi (L. Viitasaari)

Received: 6 November 2024, Revised: 12 March 2025, Accepted: 16 May 2025,
Published online: 17 June 2025

Abstract The projected normal distribution, with isotropic variance, on the 2-sphere is considered using intrinsic statistics. It is shown that in this case, the expectation commutes with the projection, and that the covariance of the normal variable has a 1-1 correspondence with the intrinsic covariance of the projected normal distribution. This allows us to estimate, after the model identification, the parameters of the underlying normal distribution that generates the data.

Keywords Spherical statistics, projected normal, parameter estimation

2010 MSC 62H11, 62F10

1 Introduction

Directional statistics (aka. spherical statistics) form a relevant subfield of statistics in which one studies directions, rotations, and axes. A typical situation in spherical statistics is that observations are gathered on a sphere, say \mathbb{S}^2 , and consequently the methods have to be adapted to non-Euclidean geometry. More generally, one can think of observations on more general compact Riemannian manifolds. Application areas are numerous as, for example, one can think of \mathbb{S}^2 representing the Earth surface, and

*Corresponding author.

measurements are then observations on this surface. To name just a few application areas, see [11, 22] for navigation/control in robotics, [14, 16] for modeling wind directional changes, [6] for finding lymphoblastic leukemia cells, [3] for movement of tectonic plates, [12] for modelling of protein chains, and [9, 19, 21] for radiology applications in the context of MIMO-systems. For details on spherical statistics, we refer to the monograph [5].

One of the central problems in statistics is to estimate model parameters from observations. In the spherical context, this can mean, for example, that one assumes a parametric distribution P_θ on \mathbb{S}^2 and then uses the data to estimate the unknown θ . The most commonly applied distributions on the sphere are von Mises–Fisher distributions (also called von Mises distributions in \mathbb{S} or Kent distributions in \mathbb{S}^2) that can be viewed as the equivalent of the normal distribution on the sphere. For the parameter estimation for von Mises–Fisher distributions, see [23]. Another widely applied distribution on \mathbb{S}^d is the projected normal distribution. That is, the distribution of $X/\|X\|$ for $X \sim N(\mu, \Sigma)$. While the density function of the projected normal is well known (see [7]), the parameter estimation is much less studied due to the complicated nature of the density. In particular, the parameter estimation is studied in the circular case (\mathbb{S}), see [15, 16, 20, 24].

In this article, we consider the problem of parameter estimation related to the projected normal distribution onto \mathbb{S}^2 . In contrast to the existing literature, we do not consider estimation of the parameters of the projected normal that can be obtained via standard methodology, as the density is completely known. Instead, our aim is to extract information on the underlying normal distribution $N(\mu, \Sigma)$ (with support on the ambient Euclidean space \mathbb{R}^3) based solely on data consisting of projected points on \mathbb{S}^2 . Obviously, we immediately run into identifiability issues if we observe only $X/\|X\|$ instead of $X \sim N(\mu, \Sigma)$. First of all, it is clear that one cannot identify arbitrary shapes of Σ (for example, the distribution can be arbitrarily spread in the direction μ and this cannot be observed as all points in this direction are equally projected). For this reason, we assume, for the sake of simplicity, isotropic variance $\Sigma = \sigma^2 I_3$ and assume $X \sim N(\mu, \sigma^2 I_3)$. However, even in this case, we can only estimate the quantities $\mu/\|\mu\|$ (the direction of the location) and $\sigma^2/\|\mu\|^2$, the reason being that the distribution of the projection $\text{pr}(aX)$ is invariant on $a > 0$. This is indeed natural, as intuitively one can only estimate the direction $\mu/\|\mu\|$ from the projections (onto \mathbb{S}^2). Similarly, $N(\mu, \sigma^2 I_3)$ for larger σ located in a distant μ seems similar to the normal distribution with smaller variance but located closer, if one is only observing both distributions on the surface of a sphere. We also note that estimation of the direction $\mu/\|\mu\|$ is already well known, and we claim no originality in this respect. Instead, our main contribution is the estimation of $\lambda = \sigma^2/\|\mu\|^2$. For this, we study the covariance matrix of $X/\|X\|$ on \mathbb{S}^2 and by linking it to certain special functions and analyzing their series expansions, we show that there is a bijective mapping between λ and the covariance matrix of $X/\|X\|$ on \mathbb{S}^2 . As the latter can be estimated by using the methods of spherical statistics, we obtain a consistent estimator for λ via the inverse mapping that can be computed, e.g., via the bisection method.

The rest of the article is organized as follows. In Section 2 we present and discuss our main results. We begin with Section 2.1 where we introduce our setup and notation. After that, we discuss the convergence of sample estimators (for mean and covariance) in the context of a general Riemannian manifold. The case of the pro-

jected normal in \mathbb{S}^2 is then discussed in Section 2.3. All the proofs are postponed to Section 3.

2 Main result

In this section we present and discuss our main results. First, in Section 2.1, we shall introduce some notation and clarify some terminology used in the context of manifold-valued random variables. Some of the theory relies on differential geometric concepts which are briefly summarized in Appendix A. The convergence of sample covariances on a general compact manifold is discussed in Section 2.2, while the special case of \mathbb{S}^2 and the projected normal distribution is treated in Section 2.3.

2.1 General setting

Let $M \subset \mathbb{R}^k$ be a smooth n -dimensional manifold. Let $S \in \mathbb{R}^k$ be a dense subset and let $\text{pr} : S \rightarrow M$ be the projection onto $M \subseteq \mathbb{R}^k$, assumed smooth everywhere on S and hence almost everywhere in \mathbb{R}^k . Let (Ω, \mathbb{P}) be a probability space and consider a normally distributed (multivariate) random variable

$$X : \Omega \rightarrow \mathbb{R}^k$$

such that

$$\mathbb{P}(X \notin S) = 0.$$

Let x_1, \dots, x_L be an independently drawn sample of X . Suppose now that we observe

$$\text{pr}(x_1), \dots, \text{pr}(x_L).$$

The question is now, can one estimate the mean and covariance of this projected sample?

In order for this question to have meaning, a notion of mean and covariance is needed for manifold-valued random variables. The intrinsic mean, a.k.a. the Fréchet mean (see [17]), of an absolutely continuous random variable $X : \Omega \rightarrow M$ is defined as follows.

Definition 2.1. Let M be a Riemannian manifold with corresponding distance function dist and volume form $d\text{Vol}_M$. Moreover, suppose $p_X : M \rightarrow [0, \infty)$ is a probability density function of some absolutely continuous random variable $X : \Omega \rightarrow M$. The expected value of X is then defined by

$$\operatorname{arg\,inf}_{q \in M} \int_M \text{dist}(q, y)^2 p_X(y) d\text{Vol}_M(y) = \mathbb{E}[X]. \quad (2.1)$$

Remark. The argument in the infimum in (2.1) need not exist nor does it need to be unique. For example, consider $X \stackrel{d}{=} N(0, \sigma^2 I) \in \mathbb{R}^{n+1}$, then $\text{pr}(X)$ follows the uniform distribution on \mathbb{S}^n and all points in \mathbb{S}^n satisfy the infimum in (2.1). This is a very important difference from the case of \mathbb{R}^n -valued random variables. In \mathbb{R}^n we know that if X is absolutely continuous, integrable and square integrable, then any point $\mu \in \mathbb{R}^n$ which minimizes the least square integral

$$\int_{\mathbb{R}^n} |x - \mu|^2 p_X dx$$

is unique.

The intrinsic definition of the covariance matrix for a random variable on a Riemannian manifold is well known as well (see, e.g., [17]).

Definition 2.2. Let M be a geodesically complete Riemannian manifold, and let $\log_q : M \rightarrow T_q M$ be the natural log map (defined a.e.). Let X be an M -valued absolutely continuous random variable with the intrinsic mean $\mu = \mathbb{E}[X] \in M$. Then, the *covariance matrix* for X is the linear map $\text{Cov}(X) : T_\mu M \rightarrow T_\mu M$ defined by the integral

$$\text{Cov}(X) = \int_M \log_\mu(y) \log_\mu(y)^T p_X(y) \, d\text{Vol}_M(y).$$

2.2 Convergence of sample estimators on a compact manifold

Let y_ℓ be a sample of L independent measurements on a compact manifold M . In order to estimate the mean of such a sample, we shall utilize the discrete version of Definition 2.1. This definition follows that of [17].

Definition 2.3. Let $\{y_\ell\}_{\ell=1}^L$ be a sample of L points on a Riemannian manifold M . Then the *empirical mean* is defined as

$$\underset{q \in M}{\text{arginf}} \sum_{\ell=1}^L \text{dist}^2(y_\ell, q).$$

Note that since the function dist^2 has good regularity, one may hope to find such an infimum by solving the (nonlinear) equation

$$q : \sum_{\ell=1}^L \log_q(y_\ell) = 0.$$

It was shown in [8] that if M is compact and if y_ℓ is sampled from a distribution with a unique empirical mean x_0 , then the distribution of the above follows a central limit theorem. More precisely, if $\mu = \mathbb{E}[y_\ell]$ is the true mean of the underlying distribution and if $\hat{\mu}$ is the empirical mean, it holds that

$$\sqrt{L} \log_\mu(\hat{\mu}) \xrightarrow{d} N(0, V)$$

for some linear map $V \in \mathcal{L}(T_\mu M, T_\mu M)$.

Definition 2.4. Let $\{y_\ell\}_{\ell=1}^L$ be a sample of L points on a Riemannian manifold M with a unique empirical mean $\hat{\xi}$. Then the *empirical covariance* of the sample is defined by

$$\hat{V} = \frac{1}{L-1} \sum_{\ell=1}^L \log_{\hat{\xi}}(y_\ell) \log_{\hat{\xi}}(y_\ell)^T \quad (2.2)$$

where $\hat{V} : T_{\hat{\xi}} M \rightarrow T_{\hat{\xi}} M$ is a linear map.

Since the empirical mean converges by the results of [8], it remains to verify that the empirical covariance in Equation (2.2) converges to the covariance V of the limiting distribution $\lim_{L \rightarrow \infty} \sqrt{L} \log_{\hat{\xi}}(\hat{\xi})$.

The following result shows that, in the case of isotropic covariance, the empirical covariance converges. The proof is postponed to Section 3.1.

Theorem 2.5. Let $\{\xi_\ell\}_{\ell=1}^L$ be L independent identically distributed random variables on a compact geodesically complete manifold M . Suppose further they have a unique mean $\mathbb{E}[\xi_\ell] = \mu$ and an isotropic covariance $\text{Cov}(\xi_\ell) = \nu I$. Then,

$$\frac{1}{L-1} \sum_{\ell=1}^L \log_{\hat{\mu}}(\xi_\ell) \log_{\hat{\mu}}(\xi_\ell)^T \xrightarrow{\mathbb{P}} \nu I$$

with the same rate of convergence as the empirical mean $\hat{\mu}$ of the sample $\{\xi_\ell\}_{\ell=1}^L$.

Remark. As far as we know, the rate of convergence for the empirical mean on compact manifolds is not completely solved as of now. In [8] it has been shown that the empirical mean has a rate of convergence \sqrt{L} for a large class of manifolds, but not all compact geodesically complete manifolds. However, [2] provides rates of convergence for the empirical mean for a general class of metric spaces, from which it seems that the rate of convergence \sqrt{L} isn't necessarily true.

Following a similar method as in [18] the update scheme for a sample of observations is given in Algorithm 1.

Algorithm 1 Estimating intrinsic average and covariance for a sample y_ℓ on a manifold M

1: Make initial guess $\hat{\mu}_0$ for average point, take, e.g., $\hat{\mu}_0 = \xi_1$.

2: Compute

$$X = \frac{1}{L} \sum_{\ell=1}^L \log_{\hat{\mu}_0}(y_\ell).$$

3: Update estimate by setting

$$\hat{\mu} = \exp_{\hat{\mu}_0}(X).$$

4: Repeat step 2-3 with $\hat{\mu}_0 = \hat{\mu}$ until $\|X\|$ is small.

5: Compute sample covariance by

$$\hat{\nu} = \frac{1}{L-1} \sum_{\ell=1}^L \log_{\hat{\mu}}(y_\ell) \log_{\hat{\mu}}(y_\ell)^T.$$

2.3 Observing the projected normal in \mathbb{S}^2

In this section we consider $M = \mathbb{S}^2$, the unit 2-sphere in \mathbb{R}^3 . In this case, the projection map is simply

$$\text{pr}(x) = \frac{x}{\|x\|}$$

and the domain of definition for pr is $S = \mathbb{R}^3 \setminus \{0\}$.

Throughout, we shall use spherical coordinates, i.e. for the classical Cartesian coordinates $(x, y, z)^T$ in \mathbb{R}^3 we set $x = \cos(\theta) \sin(\phi)$, $y = \sin(\theta) \sin(\phi)$, and $z = \cos(\phi)$. Consider two points $(\theta_1, \phi_1)^T$ and $(\theta_2, \phi_2)^T$. Then it is a classical result (see, e.g., [10]) that the distance function may be written as

$$\begin{aligned} \text{dist}((\theta_1, \phi_1)^T, (\theta_2, \phi_2)^T) \\ = \text{acos}(\cos(\phi_1) \cos(\phi_2) + \sin(\phi_1) \sin(\phi_2) \cos(\theta_2 - \theta_1)). \end{aligned}$$

It has been shown in [7] that the projected normal, i.e. the random variable defined by $\text{pr}(X)$ where $X \stackrel{d}{=} N(\mu, \Sigma)$, has the probability density function

$$p_{\text{pr}(X)}(\theta, \phi) = \left(\frac{1}{2\pi A} \right)^{3/2} |\Sigma|^{-\frac{1}{2}} \exp(C) \left(K + K^2 \frac{\Phi(K)}{\varphi(K)} + \frac{\Phi(K)}{\varphi(K)} \right) \quad (2.3)$$

at a point $u = (\cos(\theta) \sin(\phi), \sin(\theta) \sin(\phi), \cos(\phi))^T$, where $A = u^T \Sigma^{-1} u$, $B = u^T \Sigma^{-1} \mu$, $C = -\frac{1}{2} \mu^T \Sigma^{-1} \mu$, and $K = BA^{-\frac{1}{2}}$. Moreover, the functions $\varphi, \Phi : \mathbb{R} \rightarrow \mathbb{R}$ are defined as

$$\varphi(x) = \exp\left(-\frac{x^2}{2}\right)$$

and

$$\Phi(a) = \int_{-\infty}^a \varphi(x) dx,$$

respectively.

In our context, this random variable is then *observed* in \mathbb{S}^2 . As \mathbb{S}^2 is an isometrically embedded Riemannian manifold, Definition 2.1 applies for the projected normal. Intuitively, the expectation of a projected normal distribution ought to be $\text{pr}(\mu)$. Unfortunately, for the fully general case of the projected normal, this conjecture will be false. However, by imposing an isotropic covariance Σ onto X we shall show that this intuition is indeed correct. This is the topic of the next result whose proof is postponed in Section 3.2.

Theorem 2.6. *Let X be a normally distributed random variable with average $\mu \in \mathbb{R}^3$ and covariance matrix Σ , i.e. $X \stackrel{d}{=} N(\mu, \Sigma) \in \mathbb{R}^3$. If $\Sigma = \sigma^2 I_3$, then $\mathbb{E}[\text{pr}(X)] = \text{pr}(\mu)$.*

Remark. The above theorem is true whenever μ is an eigenvector of Σ . However, if μ is not an eigenvector of Σ , then the statement of Theorem 2.6 is false in general. To see this, let $\mu = (0, 0, 1)^T$ and

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}.$$

In this case it follows that the probability density function is not symmetric around μ , see Figure 1 below.

In general, the tangent space of \mathbb{S}^2 at a point q is the set of vectors orthogonal (in the Euclidean product inherited from \mathbb{R}^3) to q . Note that the initial velocity of the geodesic connecting a point μ to q has the same direction as $q \times \mu$ (i.e. the cross-product of \mathbb{R}^3). The logarithm map of q centered at a point μ is thus the vector of the normalised starting velocity of the geodesic connecting μ with q times the length of the geodesic connecting μ and q , see [13]. Without loss of generality, consider $\mu = (0, 0, 1)^T$. Then

$$\log_{\mathbb{S}^2, \mu}(q) = \begin{pmatrix} \cos(\theta) \sin(\phi) \\ \sin(\theta) \sin(\phi) \end{pmatrix} \frac{\phi}{|\sin(\phi)|},$$

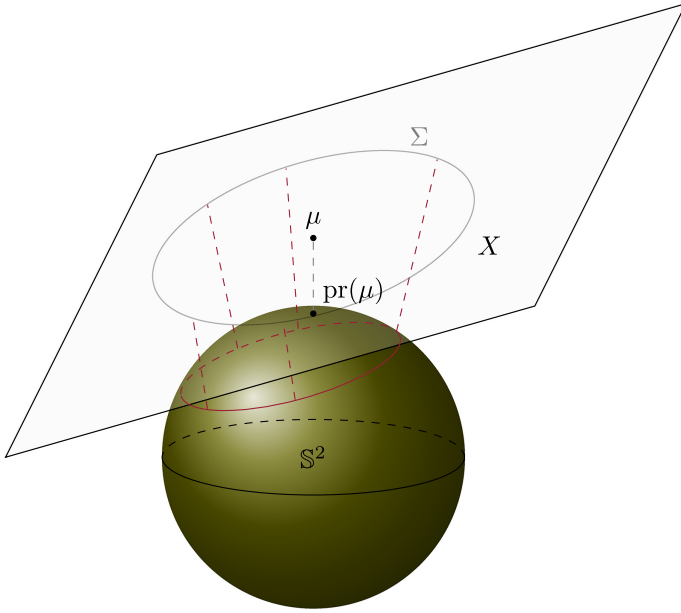


Fig. 1. A normally distributed random variable $X \stackrel{d}{=} N(\mu, \Sigma)$, with the covariance matrix Σ such that the mean μ is not an eigenvector for Σ , for which $\mathbb{E}[\text{pr}(X)] \neq \text{pr}(\mu)$

where the basis for the vector is in a rewritten form of the basis (of $T_\mu \mathbb{S}^2$)

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Hence the *intrinsic* covariance of $\text{pr}(X)$ can be written as

$$\begin{aligned} \text{Cov}(\text{pr}(X)) &= \int_0^{2\pi} \int_0^\pi \begin{pmatrix} \cos^2(\theta) \sin^2(\phi) & \cos(\theta) \sin(\theta) \sin^2(\phi) \\ \cos(\theta) \sin(\theta) \sin^2(\phi) & \sin^2(\theta) \sin^2(\phi) \end{pmatrix} \\ &\quad \times \frac{\phi^2}{\sin^2(\phi)} \left(\frac{1}{2\pi} \right)^{3/2} \exp\left(-\frac{1}{2\sigma^2}\right) \\ &\quad \times \left(\mathbf{K} + \mathbf{K}^2 \frac{\Phi(\mathbf{K})}{\varphi(\mathbf{K})} + \frac{\Phi(\mathbf{K})}{\varphi(\mathbf{K})} \right) \sin(\phi) d\phi d\theta, \end{aligned}$$

where $\mathbf{K} = \frac{1}{\sigma} \cos(\phi)$. Or, equivalently, by simplifying and integrating w.r.t. θ ,

$$\begin{aligned} \text{Cov}(\text{pr}(X)) &= \frac{\pi}{(2\pi)^{3/2}} \exp\left(-\frac{1}{2\sigma^2}\right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &\quad \times \int_0^\pi \phi^2 \left(\frac{\cos(\phi)}{\sigma} + \left(\frac{\cos^2(\phi)}{\sigma^2} + 1 \right) \frac{\Phi\left(\frac{\cos(\phi)}{\sigma}\right)}{\varphi\left(\frac{\cos(\phi)}{\sigma}\right)} \right) \sin(\phi) d\phi \quad (2.4) \\ &=: \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} f(\sigma). \end{aligned}$$

Remark. In the limit as $\sigma \rightarrow \infty$, it holds that $p_{\text{pr}(X)} \rightarrow \frac{1}{4\pi} \mathbb{1}_{\mathbb{S}^2}$. The intrinsic covariance in this limit is

$$\frac{\pi^2 - 4}{4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

On the other hand, in the limit as $\sigma \rightarrow 0$, $p_{\text{pr}(X)}$ converges to the point mass distribution while the covariance goes to zero.

Note that the above estimates are only for the *intrinsic* expectation and covariance. If we want to relate these estimates to the *extrinsic* expectation and covariance, we end up with the obvious problem that $\text{pr}(X)$ and $\text{pr}(aX)$, $a > 0$, have the same distribution. Hence we need to choose which normal distribution in \mathbb{R}^3 corresponds to the intrinsic covariance and average. By Lemma 2.6 the average is known (up to a factor). Moreover, it turns out there is a one-to-one correspondence between the extrinsic covariance and the intrinsic covariance up to a factor $a > 0$, and thus one may estimate the underlying covariance parameter σ using the relation (2.4) with, e.g., the bisection method.

The degeneracy of the factor a essentially means that we can only estimate the parameter σ in the case when the projected average is the *true* average of the \mathbb{R}^3 -valued normal random variable. More generally, if $X \stackrel{\text{d}}{=} N(\mu, \sigma^2 I_3)$, then $\frac{X}{\|\mu\|} \stackrel{\text{d}}{=} N\left(\frac{\mu}{\|\mu\|}, \frac{\sigma^2}{\|\mu\|^2} I_3\right)$. Consequently, we can only estimate the quantities $\frac{\mu}{\|\mu\|} \in \mathbb{S}^2$ and $\frac{\sigma^2}{\|\mu\|^2}$. This means that, as expected, we can estimate the direction $\frac{\mu}{\|\mu\|}$ of the extrinsic distribution $X \stackrel{\text{d}}{=} N(\mu, \sigma^2 I_3)$ but not the distance $\|\mu\|$. On the other hand, for the variance we can only estimate the quantity $\frac{\sigma^2}{\|\mu\|^2}$. This is expected as well, as for the distribution $N(\mu, \sigma^2 I_3)$ located far away ($\|\mu\|$ large) the projections onto \mathbb{S}^2 are not wide spread even if $\sigma > 0$ is large.

The required one-to-one correspondence between extrinsic and intrinsic covariances is formulated in the following proposition, whose proof is presented in Section 3.3. The statement is also illustrated in Figure 2.

Proposition 2.7. *Let X be a normally distributed random variable with mean $\mu \in \mathbb{R}^3$ and isotropic variance $\sigma^2 I_3$, i.e. $\frac{X}{\|\mu\|} \stackrel{\text{d}}{=} N\left(\frac{\mu}{\|\mu\|}, \frac{\sigma^2}{\|\mu\|^2} I_3\right)$. Then*

$$\text{Cov}(\text{pr}(X)) = v I_2$$

for $0 \leq v < \frac{\pi^2 - 4}{4}$ and the relation

$$f\left(\frac{\sigma^2}{\|\mu\|^2}\right) := \frac{\text{tr}(\text{Cov}(\text{pr}(X)))}{2} = v$$

is a bijection.

The bijection f is crucial for observing projected normal random variables. Utilizing this f , if the scalar variance of some normal random variable $X \stackrel{\text{d}}{=} N(\mu, \sigma^2 I_3)$ in \mathbb{R}^3 is known, then the intrinsic scalar variance of the corresponding projected normal random variable is precisely $f\left(\frac{\sigma^2}{\|\mu\|^2}\right)$. Conversely, if some projected isotropic normal random variable has mean μ and covariance $v I_2$, then the normal random variable X

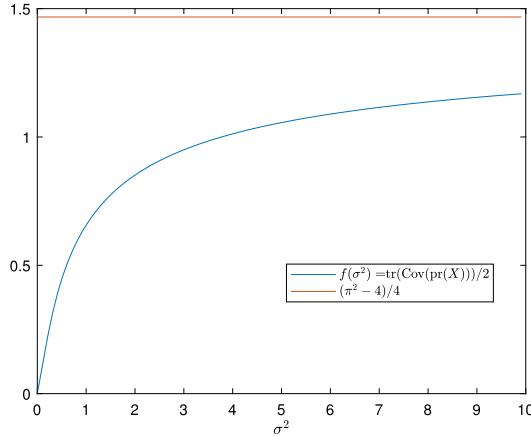


Fig. 2. A plot of the scalar variance of $\text{pr}(X)$, i.e. $f(\sigma^2)$ from Proposition 2.7, where $X \stackrel{\text{d}}{=} N(\mu, \sigma^2 I_3)$ and $\mu \in \mathbb{S}^2$ is arbitrary. The red line indicates the upper bound $(\pi^2 - 4)/4 = \lim_{\sigma \rightarrow \infty} \frac{\text{tr}(\text{Cov}(\text{pr}(X)))}{2}$

such that $\text{pr}(X)$ has these parameters is precisely $X \stackrel{\text{d}}{=} N(\mu, f^{-1}(v)I_3)$ (up to a positive scalar factor).

As a consequence of Proposition 2.7 and Theorem 2.5 we can conclude that given measurements on the sphere, we can estimate the scalar variance of an isotropically distributed normal random vector in \mathbb{R}^3 . This leads to the next result that can be viewed as the main theorem of the present paper. Its proof follows directly from Theorem 2.5 and Proposition 2.7. The rate of convergence is obtained immediately from noting that [8, Theorem 2] applies to \mathbb{S}^2 , and thus the empirical mean has rate of convergence \sqrt{L} , and by Theorem 2.5 so does the empirical covariance.

Theorem 2.8. *Let X be a normally distributed random variable with mean $\mu \in \mathbb{R}^3$ and isotropic variance $\sigma^2 I_3$, i.e. $X \stackrel{\text{d}}{=} N(\mu, \sigma^2 I_3)$. Given independent measurements (x_1, x_2, \dots, x_L) from $\text{pr}(X)$, we can estimate $\lambda = \frac{\sigma^2}{\|\mu\|^2}$ by*

$$\hat{\lambda} = f^{-1} \left(\frac{\text{tr}(\hat{V})}{2} \right)$$

where \hat{V} is the empirical covariance matrix given in Equation (2.2) and where f is the bijection given in Proposition 2.7 and defined in Equation (2.4). Moreover, it holds that

$$\hat{\lambda} \xrightarrow{\mathbb{P}} \frac{\sigma^2}{\|\mu\|^2}$$

as $L \rightarrow \infty$, with rate of convergence \sqrt{L} .

We conclude this section with some simulations. Firstly, Table 1 illustrates an empirical verification of Theorem 2.5 for the case of the manifold \mathbb{S}^2 . The simulations are conducted by generating L samples from $X \stackrel{\text{d}}{=} N(\mu, \sigma^2 I_3)$ with $\mu(0, 0, 1)^T \in \mathbb{S}^2$

Table 1. The absolute error of 100 times repeated Monte Carlo simulations of the empirical covariance, Equation (2.2), compared to the theoretical covariance of the projected normal distribution, $\text{tr}(\text{Cov}(\text{pr}(X)))/2$. Each simulation run uses L data points and $\sigma = 1$

L	30	50	100	1000	10^4	10^5	10^6
error	0.013	0.0080	0.0055	0.0033	0.0015	8.4e-05	2.9e-05

and $\sigma = 1$ and projecting these samples onto \mathbb{S}^2 . The sample covariance is computed as in Algorithm 1, and its error is then compared to the theoretical covariance using the Frobenius norm. This error is then averaged over using 100 repetitions of the Monte Carlo method.

Secondly, Figure 3 is an illustration of the convergence result in Theorem 2.8. The simulations are done by generating L samples from $X \stackrel{d}{=} N(\mu, \sigma^2 I_3)$ with $\mu(0, 0, 1)^T \in \mathbb{S}^2$ and $\sigma = 1$. The empirical covariance \hat{V} , Equation (2.2), is computed for each set of samples. Then, for each set of samples, the estimator $\hat{\lambda}$ for σ^2 , see Theorem 2.8, is computed by $f^{-1}(\text{tr}(\hat{V})/2)$. This is done with a 1000-fold repetition.

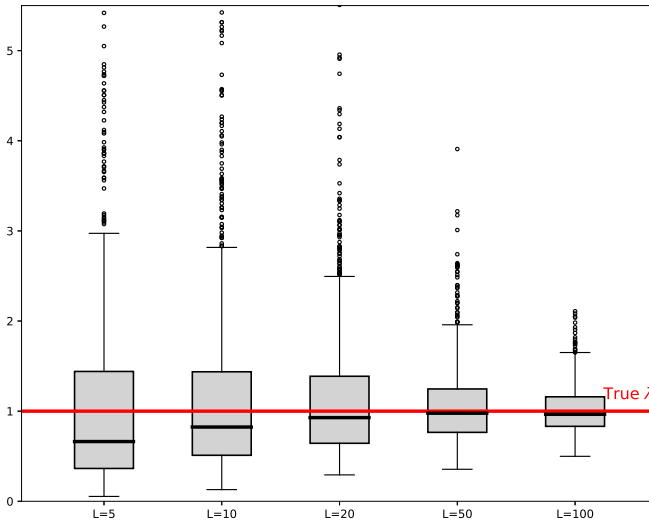


Fig. 3. A box plot of the inferred estimator $\hat{\lambda}$ in Theorem 2.8 using the empirical covariance, Equation (2.2), for L randomly generated projected normal random variables with $\sigma^2 = 1$ and $\mu = (0, 0, 1)^T$. For each L this is done with 1000 repetitions. The true underlying scalar variance, $\lambda = f^{-1}(\text{tr}(\text{Cov}(\text{pr}(X)))/2) = 1$, is shown in red color. Note that some outliers are omitted for readability of the graph

3 Proofs

3.1 Proof of Theorem 2.5

First, note that

$$\frac{1}{L-1} \sum_{\ell=1}^L \log_{\hat{\mu}}(\xi_{\ell}) \log_{\hat{\mu}}(\xi_{\ell})^T$$

is a linear map $T_{\hat{\mu}}\mathbb{S}^2 \rightarrow T_{\hat{\mu}}\mathbb{S}^2$, and $\text{Cov}(\xi)$ is by definition a linear map $T_{\mu}\mathbb{S}^2 \rightarrow T_{\mu}\mathbb{S}^2$. In order to compare \hat{V} and $\text{Cov}(\xi)$ we shall transport parallelly the linear map $\hat{V} : T_{\hat{\mu}} \rightarrow T_{\hat{\mu}}$ to a linear map $T_{\mu}\mathbb{S}^2 \rightarrow T_{\mu}\mathbb{S}^2$. That is, we look at $P_{\hat{\mu},\mu} \hat{V} P_{\mu,\hat{\mu}}$ and see how far away it is from vI_2 . It follows from Equation (A.1) that $P_{\hat{\mu},\mu} \log_{\hat{\mu}}(\xi_\ell) \log_{\hat{\mu}}(\xi_\ell)^T P_{\mu,\hat{\mu}}$ has the first order expansion

$$\begin{aligned} & P_{\hat{\mu},\mu} \log_{\hat{\mu}}(\xi_\ell) \log_{\hat{\mu}}(\xi_\ell)^T P_{\mu,\hat{\mu}} \\ &= \log_{\mu}(\xi_\ell) \log_{\mu}(\xi_\ell)^T + \left(\nabla_{\log_{\mu}(\hat{\mu})} \log_{\mu}(\xi_\ell) \right) \log_{\mu}(\xi_\ell)^T \\ & \quad + \log_{\mu}(\xi_\ell) \left(\nabla_{\log_{\mu}(\hat{\mu})} \log_{\mu}(\xi_\ell) \right)^T + \mathcal{O}(\text{dist}(\mu, \hat{\mu})^2). \end{aligned}$$

Therefore,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left\| \mathbb{E} \left[P_{\hat{\mu},\mu} \frac{1}{n-1} \sum_{\ell=1}^n \log_{\hat{\mu}}(\xi_\ell) \log_{\hat{\mu}}(\xi_\ell)^T P_{\mu,\hat{\mu}} - vI \right] \right\| \\ & \leq \lim_{n \rightarrow \infty} \left\| \frac{1}{n-1} \sum_{\ell=1}^n \mathbb{E} \left[\log_{\mu}(\xi_\ell) \log_{\mu}(\xi_\ell)^T \right] - vI \right\| \\ & \quad + \lim_{n \rightarrow \infty} \left\| \mathbb{E} \left[\left(\nabla_{\log_{\mu}(\hat{\mu})} \log_{\mu}(\xi_\ell) \right) \log_{\mu}(\xi_\ell)^T + \log_{\mu}(\xi_\ell) \left(\nabla_{\log_{\mu}(\hat{\mu})} \log_{\mu}(\xi_\ell) \right)^T \right] \right\| \\ & \quad + \lim_{n \rightarrow \infty} \mathbb{E} \left[\mathcal{O}(\text{dist}(\hat{\mu}, \mu)^2) \right] \\ & \leq \|vI - vI\| + \lim_{n \rightarrow \infty} \mathbb{E} \left[\mathcal{O}(\text{dist}(\hat{\mu}, \mu)) \right] \end{aligned}$$

Here, according to [8, Proposition 1], it holds that $\text{dist}(\mu, \hat{\mu})$ converges to zero almost surely. Since M is compact, it has finite mass and thus $\text{dist}(\mu, \hat{\mu})$ converges in expectation to 0 by the dominated convergence theorem. Finally, the fact that the empirical covariance has the same rate of convergence as the empirical mean follows from the last inequality. This completes the proof. \square

3.2 Proof of Theorem 2.6

By the very definition, we need to show that

$$\text{arginf}_{q \in \mathbb{S}^2} \int_{\mathbb{S}^2} \text{dist}(q, y)^2 p_{\text{pr}(X)}(y) d\mathbb{S}^2(y) = \text{pr}(\mathbb{E}[X]) \quad (3.1)$$

for $X \stackrel{d}{=} N(\mu, \sigma^2 I_3)$. By rescaling X with a factor of $\frac{1}{\|\mu\|}$ and by a rotational symmetry, we may assume $\mu = (-1, 0, 0)^T$ without loss of generality.

We make a simple first-derivative test to find the minimum in (3.1). Let (θ_1, ϕ_1) be the spherical coordinates that are integrated over the sphere and let (θ_2, ϕ_2) be the coordinates that minimize the integral. Hence we need to show that $(\theta_2, \phi_2) = (\pi, \pi/2)$. Writing the integral in spherical coordinates and using the fact that the derivative with respect to ϕ_2 is zero, it follows that (θ_2, ϕ_2) satisfies

$$\frac{\partial}{\partial \phi_2} \int_{\mathbb{S}^2} \text{dist}^2((\theta_1, \phi_1), (\theta_2, \phi_2)) p_{\text{pr}(X)}(\theta_1, \phi_1) \sin(\phi_1) d\theta_1 d\phi_1 = 0.$$

Now by the Leibniz integral rule the derivative may be moved inside the integral, and it holds

$$\begin{aligned} & \int_0^{2\pi} \int_0^\pi 2 \operatorname{dist}((\theta_1, \phi_1), (\theta_2, \phi_2)) \\ & \quad \times \frac{(\cos(\phi_1) \sin(\phi_2) - \sin(\phi_1) \cos(\phi_2) \cos(\theta_2 - \theta_1))}{\sqrt{1 - (\cos(\phi_1) \cos(\phi_2) + \sin(\phi_1) \sin(\phi_2) \cos(\theta_2 - \theta_1))^2}} \\ & \quad \times p_{\operatorname{pr}(X)}(\theta_1, \phi_1) \sin(\phi_1) d\theta_1 d\phi_1 \\ & = 0. \end{aligned}$$

Plugging $(\theta_2, \phi_2) = (\pi, \pi/2)$ into the integral on the left-hand side yields an integral

$$\begin{aligned} & 2 \int_0^{2\pi} \int_0^\pi \operatorname{acos}(\sin(\phi_1) \cos(\pi - \theta_1)) \frac{\cos(\phi_1)}{\sqrt{1 - \sin^2(\phi_1) \cos^2(\pi - \theta_1)}} \\ & \quad \times p_{\operatorname{pr}(X)}(\theta_1, \phi_1) \sin(\phi_1) d\theta_1 d\phi_1. \end{aligned}$$

This integral equals zero, which can be seen by observing that the integrand is odd along ϕ_1 around $\frac{\pi}{2}$, since $p_{\operatorname{pr}(X)}$ is symmetric around μ by construction. By similar arguments, we obtain that

$$\frac{\partial}{\partial \theta_2} \int_{\mathbb{S}^2} \operatorname{dist}^2((\theta_1, \phi_1), (\theta_2, \phi_2)) p_{\operatorname{pr}(X)}(\theta_1, \phi_1) \sin(\phi_1) d\phi_1 d\theta_1$$

reduces at $(\theta_2, \phi_2) = (\pi, \pi/2)$ to

$$\begin{aligned} & 2 \int_0^{2\pi} \int_0^\pi \operatorname{acos}(\sin(\phi_1) \cos(\pi - \theta_1)) \frac{\sin^2(\phi_1) \sin(\pi - \theta_1)}{\sqrt{1 - \sin^2(\phi_1) \cos^2(\pi - \theta_1)}} \\ & \quad \times p_{\operatorname{pr}(X)}(\theta_1, \phi_1) d\phi_1 d\theta_1 \end{aligned}$$

which equals zero by symmetry around $\theta_1 = \pi$. Therefore, one can conclude that $\mu = \operatorname{pr}(\mathbb{E}[X])$ is a local extremum for the integral in Equation (3.1).

Next we shall argue why it is a global minimum. Note that a very similar computation will show that $-\mu$, i.e. in spherical coordinates $(\theta_2, \phi_2) = (0, \pi/2)$ is another local extremum. In fact, μ and $-\mu$ are the only local extremum for the integral in Equation (3.1) since it is the only two points for which the distance function is rotationally symmetric around the line in \mathbb{R}^3 which is spanned by μ . By direct computation,

$$p_{\operatorname{pr}(X)}\left(\pi, \frac{\pi}{2}\right) = \frac{1}{(2\pi)^{3/2}} \exp\left(-\frac{1}{2\sigma^2}\right) \left(\frac{1}{\sigma} + \frac{1}{\sigma^2} \frac{\Phi(\frac{1}{\sigma})}{\varphi(\frac{1}{\sigma})} + \frac{\Phi(\frac{1}{\sigma})}{\varphi(\frac{1}{\sigma})}\right)$$

and

$$p_{\operatorname{pr}(X)}\left(0, \frac{\pi}{2}\right) = \frac{1}{(2\pi)^{3/2}} \exp\left(-\frac{1}{2\sigma^2}\right) \left(\frac{-1}{\sigma} + \frac{1}{\sigma^2} \frac{\Phi(\frac{-1}{\sigma})}{\varphi(\frac{-1}{\sigma})} + \frac{\Phi(\frac{-1}{\sigma})}{\varphi(\frac{-1}{\sigma})}\right),$$

hence $p_{\operatorname{pr}(X)}(\mu) > p_{\operatorname{pr}(X)}(-\mu)$. More generally, a similar computation shows that $p_{\operatorname{pr}(X)}(y) > p_{\operatorname{pr}(X)}(Ry)$, if $\langle y, \mu \rangle > 0$, where R is the reflection mapping over the

y, z -plane, i.e.

$$R \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -a \\ b \\ c \end{pmatrix}.$$

Tautologically, the distance function increases when one goes father, and points being farther away will contribute more into the integral. Therefore $(\theta_2, \phi_2) = (\pi, \pi/2)$ yields a global minimum for the integral in Equation (3.1).

□

3.3 Proof of Proposition 2.7

Without loss of generality, we assume $\mu = (0, 0, 1)^T$. In this case the functions A, B, C, K inside Equation (2.3) are given by

$$\begin{aligned} A &= \frac{1}{\sigma^2}, \\ B &= \frac{1}{\sigma^2} \cos(\phi), \\ C &= -\frac{1}{2\sigma^2}, \end{aligned}$$

and

$$K = \frac{1}{\sigma} \cos(\phi).$$

By (2.4) it holds that $\text{Cov}(\text{pr}(X))$ is isotropic and we can write

$$\begin{aligned} \frac{\text{tr}(\text{Cov}(\text{pr}(X)))}{2} &= \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}\right) \\ &\times \int_0^\pi \phi^2 \left(\frac{\cos(\phi)}{\sigma} + \left(\frac{\cos^2(\phi)}{\sigma^2} + 1 \right) \frac{\Phi\left(\frac{\cos(\phi)}{\sigma}\right)}{\varphi\left(\frac{\cos(\phi)}{\sigma}\right)} \right) \sin(\phi) d\phi. \end{aligned}$$

Denote

$$f(\sigma) = \exp\left(-\frac{1}{2\sigma^2}\right) \int_0^\pi \phi^2 \left(\frac{\cos(\phi)}{\sigma} + \left(\frac{\cos^2(\phi)}{\sigma^2} + 1 \right) \frac{\Phi\left(\frac{\cos(\phi)}{\sigma}\right)}{\varphi\left(\frac{\cos(\phi)}{\sigma}\right)} \right) \sin(\phi) d\phi.$$

In order to obtain the claim, it suffices to prove that $f(\sigma)$ is strictly increasing, from which it follows that $\frac{\text{tr}(\text{Cov}(\text{pr}(X)))}{2}$ is strictly increasing in σ as well. For notational simplicity, we set $x = \frac{1}{\sigma}$ and show that $f(x)$ is strictly decreasing in x , where now $f(x)$, with a slight abuse of notation, is given by

$$\begin{aligned} f(x) &= \exp\left(-\frac{x^2}{2}\right) \\ &\times \int_0^\pi \phi^2 \left(x \cos(\phi) + (x^2 \cos^2(\phi) + 1) \frac{\Phi(x \cos(\phi))}{\varphi(x \cos(\phi))} \right) \sin(\phi) d\phi. \end{aligned} \tag{3.2}$$

By differentiating in x , we get

$$f'(x) = \exp\left(\frac{-x^2}{2}\right)(I_1 + I_2 + I_3), \quad \text{where}$$

$$I_1 := -x \int_0^\pi \phi^2 \cos(\phi) x \sin(\phi) d\phi,$$

$$I_2 := \int_0^\pi \phi^2 (-x^3 \cos^2(\phi) - x + 3x \cos^2(\phi) + x^3 \cos^4(\phi)) \frac{\Phi(x \cos(\phi))}{\varphi(x \cos(\phi))} \sin(\phi) d\phi,$$

$$I_3 := \int_0^\pi \phi^2 (2 \cos(\phi) + x^2 \cos^3(\phi)) \sin(\phi) d\phi.$$

Immediately, we have that

$$I_1 = x^2 \frac{\pi^2}{4}$$

and

$$I_3 = -\frac{\pi^2}{2} - \frac{5\pi^2}{32} x^2.$$

In order to show $f'(x) < 0$, we need to decipher I_2 , which is more complicated. The main idea is to show that $f'(x)$ is analytic for all $x \geq 0$ and then to show that there is a strictly decreasing analytic function between $f'(x)$ and 0. We begin with Lemma 3.1 showing that $\frac{\Phi}{\varphi}$ is a Dawson-like function and therefore analytic. In Lemma 3.2 the series expansion of I_2 is integrated term-wise. The terms of the series expression for I_2 are given inductively in Lemma 3.3, which are then inserted to give explicit expressions for the terms of $f'(x) \exp(x^2/2)$ in Lemma 3.4. Lemma 3.5 then gives a lower bound on the odd terms of $f'(x) \exp(x^2/2)$, and Lemma 3.7 gives an upper bound. By Lemma 3.8 we show that the series expression of $f'(x) \exp(x^2/2)$ is eventually decreasing as a series, and in combination with Corollary 3.6 it is concluded that $f'(x) \exp(x^2/2)$ is entire. The proof is finished by comparing $f'(x) \exp(x^2/2)$ to a linear combination of analytic functions computed in Lemma 3.9.

Lemma 3.1. *It holds that*

$$\frac{\Phi}{\varphi}(x) = \sum_{k=0}^{\infty} \frac{1}{(2k+1)!!} x^{2k+1} + \frac{\sqrt{2\pi}}{2} \sum_{k=0}^{\infty} \frac{1}{(2k)!!} x^{2k} =: \sum_{k=0}^{\infty} d_k x^k$$

where the series converges everywhere.

Proof. Note that $\frac{\Phi}{\varphi}(x)$ is very similar to the Dawson function (originally studied in [4], see also [1] for further details) given by

$$D_-(x) = \exp(x^2) \int_0^x \exp(-t^2) dt,$$

and it has the series expansion

$$D_-(x) = \sum_{k=0}^{\infty} \frac{2^k}{(2k+1)!!} x^{2k+1}.$$

Note that

$$\int_0^x \exp\left(-\frac{t^2}{2}\right) dt = \sqrt{2} \int_0^{x/\sqrt{2}} \exp(-u^2) du$$

by the variable substitution $u = t/\sqrt{2}$. Hence,

$$\begin{aligned} \exp\left(\frac{x^2}{2}\right) \int_0^x \exp\left(-\frac{t^2}{2}\right) dt &= \sqrt{2} D_-(x/\sqrt{2}) \\ &= \sqrt{2} \sum_{k=0}^{\infty} \frac{2^k}{(2k+1)!!} \frac{x^{2k+1}}{2^k \sqrt{2}} = \sum_{k=0}^{\infty} \frac{1}{(2k+1)!!} x^{2k+1}. \end{aligned}$$

Moreover,

$$\int_{-\infty}^0 \exp\left(-\frac{t^2}{2}\right) dt = \frac{\sqrt{2\pi}}{2}$$

and hence

$$\exp\left(\frac{x^2}{2}\right) \int_{-\infty}^0 \exp\left(-\frac{t^2}{2}\right) dt$$

has the series expansion

$$\frac{\sqrt{2\pi}}{2} \sum_{k=0}^{\infty} \frac{1}{k!} \frac{x^{2k}}{2^k} = \frac{\sqrt{2\pi}}{2} \sum_{k=0}^{\infty} \frac{1}{(2k)!!} x^{2k}.$$

It follows that

$$\begin{aligned} \frac{\Phi}{\varphi}(x) &= \exp\left(\frac{x^2}{2}\right) \int_{-\infty}^0 \exp\left(-\frac{t^2}{2}\right) dt + \exp\left(\frac{x^2}{2}\right) \int_0^x \exp\left(-\frac{t^2}{2}\right) dt \\ &= \sum_{k=0}^{\infty} \frac{1}{(2k+1)!!} x^{2k+1} + \frac{\sqrt{2\pi}}{2} \sum_{k=0}^{\infty} \frac{1}{(2k)!!} x^{2k} \end{aligned}$$

proving the claimed series representation. Finally, the convergence everywhere follows from the fact that the Dawson function converges everywhere. \square

Using Lemma 3.1 it follows that I_2 can be rewritten as

$$\begin{aligned} I_2 &= \sum_{k=0}^{\infty} d_k \int_0^{\pi} \phi^2 (-x^3 \cos^2(\phi) - x + 3x \cos^2(\phi) + x^3 \cos^4(\phi)) \\ &\quad \times x^k \cos^k(\phi) \sin(\phi) d\phi. \end{aligned} \tag{3.3}$$

Each term in Equation (3.3) involves

$$J_m = \int_0^{\pi} \phi^2 \cos^m(\phi) \sin(\phi) d\phi$$

which we will study next.

Lemma 3.2. *The sequence J_m satisfies $J_0 = \pi^2 - 4$, and for even $m \neq 0$*

$$J_m = \frac{1}{m+1} \left(\pi^2 - 4 \frac{m!!}{(m+1)!!} \left(1 + \sum_{j=0}^{\frac{m}{2}-1} \frac{1}{2^{2j+1} (2j+3) \binom{2j+1}{j}} \right) \right),$$

while for $m \in \mathbb{N}$ odd

$$J_m = \frac{\pi^2}{m+1} \left(\frac{m!!}{(m+1)!!} - 1 \right).$$

Proof. For J_0 we observe immediately that

$$J_0 = \int_0^\pi \phi^2 \sin(\phi) d\phi = \pi^2 - 4.$$

Next, let m be odd. Then integration by parts gives

$$\begin{aligned} J_m &= \left[-\phi^2 \frac{\cos^{m+1}(\phi)}{m+1} \right]_0^\pi + \frac{2}{m+1} \int_0^\pi \phi \cos^{m+1}(\phi) d\phi \\ &= -\frac{\pi^2}{m+1} + \frac{2}{m+1} \left[\frac{m!!}{(m+1)!!} \phi^2 + \phi \sum_{j=0}^{\frac{m-1}{2}} \cos^{m-2j}(\phi) \sin(\phi) \frac{m!!}{(m-2j)!!} \right. \\ &\quad \times \left. \frac{(m-2j-1)!!}{(m+1)!!} \right]_0^\pi - \frac{2}{m+1} \sum_{j=0}^{\frac{m-1}{2}} \frac{m!!}{(m-2j)!!} \frac{(m-2j-1)!!}{(m+1)!!} \\ &\quad \times \int_0^\pi \cos^{m-2j}(\phi) \sin(\phi) d\phi - \frac{2}{m+1} \frac{m!!}{(m+1)!!} \int_0^\pi \phi d\phi \\ &= -\frac{\pi^2}{m+1} + \frac{2}{m+1} \frac{m!!}{(m+1)!!} \pi^2 - \frac{1}{m+1} \frac{m!!}{(m+1)!!} \pi^2 \\ &= \frac{\pi^2}{m+1} \left(\frac{m!!}{(m+1)!!} - 1 \right). \end{aligned}$$

Similarly, when $m > 0$ is even, integration by parts gives

$$\begin{aligned} J_m &= \left[-\phi^2 \frac{\cos^{m+1}(\phi)}{m+1} \right]_0^\pi + \frac{2}{m+1} \int_0^\pi \phi \cos^{m+1}(\phi) d\phi \\ &= \frac{\pi^2}{m+1} + \frac{2}{m+1} \left[\phi \sin(\phi) \frac{m!!}{(m+1)!!} + \phi \sum_{j=0}^{\frac{m}{2}-1} \cos^{m-2j}(\phi) \sin(\phi) \frac{m!!}{(m-2j)!!} \right. \\ &\quad \times \left. \frac{(m-1-2j)!!}{(m+1)!!} \right]_0^\pi - \frac{2}{m+1} \frac{m!!}{(m+1)!!} \int_0^\pi \sin(\phi) d\phi \\ &\quad - \frac{2}{m+1} \sum_{j=0}^{\frac{m}{2}-1} \frac{m!!}{(m-2j)!!} \frac{(m-1-2j)!!}{(m+1)!!} \int_0^\pi \cos^{m-2j}(\phi) \sin(\phi) d\phi \\ &= \frac{\pi^2}{m+1} - \frac{4}{m+1} \frac{m!!}{(m+1)!!} \end{aligned}$$

$$\begin{aligned}
& -\frac{4}{m+1} \frac{m!!}{(m+1)!!} \sum_{j=0}^{\frac{m}{2}-1} \frac{(m-1-2j)!!}{(m-2j)!!(m+1-2j)} \\
& = \frac{1}{m+1} \left(\pi^2 - 4 \frac{m!!}{(m+1)!!} \left(1 + \sum_{j=0}^{\frac{m}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} \right) \right).
\end{aligned}$$

This completes the proof. \square

Lemma 3.3. Denote

$$I_2 = \sum_{n=1}^{\infty} a_n x^n.$$

Then the coefficients a_k satisfy

$$a_1 = \frac{4\sqrt{2\pi}}{9}, \quad a_2 = -\frac{7\pi^2}{32},$$

and

$$a_k = d_{k-1}(3J_{k+1} - J_{k-1}) + d_{k-3}(J_{k+1} - J_{k-1}), \quad k \geq 3.$$

Proof. Immediately $I_2(0) = 0$, so the constant term is zero. Next, we get expressions for a_n in terms of d_n 's and J_n 's from Equation (3.3). Then straightforward calculations give

$$\begin{aligned}
a_1 & = d_0(-J_0 + 3J_2) = \frac{\sqrt{2\pi}}{2} \left(4 - \pi^2 + \frac{3}{3} \left(\pi^2 - 4 \frac{2}{3} \left(1 + \frac{1}{6} \right) \right) \right) = \frac{4\sqrt{2\pi}}{9}, \\
a_2 & = d_1(-J_1 + 3J_3) = \frac{\pi^2}{4} + \frac{3\pi^2}{4} \left(\frac{3}{8} - 1 \right) = -\frac{7\pi^2}{32},
\end{aligned}$$

and, for $k \geq 3$,

$$a_k = d_{k-1}(3J_{k+1} - J_{k-1}) + d_{k-3}(J_{k+1} - J_{k-1}).$$

\square

Lemma 3.4. Denote

$$f'(x) \exp\left(\frac{x^2}{2}\right) = \sum_{k=0}^{\infty} c_k x^k. \quad (3.4)$$

Then

$$c_0 = -\frac{\pi^2}{2}, \quad c_1 = \frac{4\sqrt{2\pi}}{9}, \quad c_2 = -\frac{\pi^2}{8},$$

for $k \geq 3$ odd

$$c_k = \frac{2\sqrt{2\pi}}{(k+2)!!} \left(1 + \sum_{j=0}^{\frac{k-1}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} - \frac{k+1}{2^k(k+2)} \binom{k}{\frac{k-1}{2}} \right),$$

and for $k \geq 4$ even

$$c_k = -\frac{\pi^2}{(k+2)!!}.$$

Proof. Considering the expressions for I_1 , I_2 , and I_3 gives

$$\begin{aligned} c_0 &= -\frac{\pi^2}{2}, \\ c_1 &= a_1 = \frac{4\sqrt{2\pi}}{9}, \\ c_2 &= a_2 + \frac{3\pi^2}{32} = -\frac{\pi^2}{8}, \end{aligned}$$

and

$$c_k = a_k \quad \text{whenever } k \geq 3,$$

where a_k is given in Lemma 3.3. It follows that for $k \geq 3$ odd we have

$$\begin{aligned} c_k &= d_{k-1}(3J_{k+1} - J_{k-1}) + d_{k-3}(J_{k+1} - J_{k-1}) \\ &= \frac{\sqrt{2\pi}}{2} \left(\frac{1}{(k-1)!!} \left[\frac{3}{k+2} \left(\pi^2 - 4 \frac{(k+1)!!}{(k+2)!!} \right. \right. \right. \\ &\quad \times \left. \left. \left. \left(1 + \sum_{j=0}^{\frac{k+1}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} \right) \right) \right] \right. \\ &\quad \left. - \frac{1}{k} \left(\pi^2 - 4 \frac{(k-1)!!}{k!!} \left(1 + \sum_{j=0}^{\frac{k-1}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} \right) \right) \right) \\ &\quad + \frac{1}{(k-3)!!} \left[\frac{1}{k+2} \left(\pi^2 - 4 \frac{(k+1)!!}{(k+2)!!} \right. \right. \\ &\quad \times \left. \left. \left. \left(1 + \sum_{j=0}^{\frac{k+1}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} \right) \right) \right] \\ &\quad \left. - \frac{1}{k} \left(\pi^2 - 4 \frac{(k-1)!!}{k!!} \left(1 + \sum_{j=0}^{\frac{k-1}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} \right) \right) \right) \right], \end{aligned}$$

and after simplifying we have

$$c_k = \frac{2\sqrt{2\pi}}{(k+2)!!} \left(1 + \sum_{j=0}^{\frac{k-1}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} - \frac{k+1}{2^k(k+2)} \binom{k}{\frac{k-1}{2}} \right).$$

Similarly for $k \geq 4$ even we get

$$\begin{aligned} c_k &= d_{k-1}(3J_{k+1} - J_{k-1}) + d_{k-3}(J_{k+1} - J_{k-1}) \\ &= \frac{\pi^2}{(k-1)!!} \left[\frac{3}{k+2} \left(\frac{(k+1)!!}{(k+2)!!} - 1 \right) - \frac{1}{k} \left(\frac{(k-1)!!}{k!!} - 1 \right) \right] \\ &\quad + \frac{\pi^2}{(k-3)!!} \left[\frac{1}{k+2} \left(\frac{(k+1)!!}{(k+2)!!} - 1 \right) - \frac{1}{k} \left(\frac{(k-1)!!}{k!!} - 1 \right) \right] \end{aligned}$$

$$= -\frac{\pi^2}{(k+2)!!}.$$

This completes the proof. \square

Lemma 3.5. For all $k \geq 3$ odd, set

$$S(k) = 1 + \sum_{j=0}^{\frac{k-1}{2}-1} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} - \frac{k+1}{2^k(k+2)} \binom{k}{\frac{k-1}{2}}. \quad (3.5)$$

Then $S(k)$ is increasing and satisfies $2\sqrt{2\pi}S(k) \geq \pi$.

Proof. First note that

$$\begin{aligned} 2\sqrt{2\pi}S(3) &= 2\sqrt{2\pi} \left(1 + \frac{1}{6} \binom{1}{0} - \frac{4}{5} \frac{1}{2^3} \binom{3}{1} \right) \\ &= 2\sqrt{2\pi} \left(1 + \frac{1}{6} - \frac{3}{10} \right) \\ &= \frac{26\sqrt{2\pi}}{15} \geq \pi. \end{aligned}$$

It remains to show that $S(k)$ is an increasing function. We have

$$\begin{aligned} S(k+2) &= 1 + \sum_{j=0}^{\frac{k-1}{2}} \frac{1}{2^{2j+1}(2j+3)} \binom{2j+1}{j} - \frac{k+3}{2^{k+2}(k+4)} \binom{k+2}{\frac{k+1}{2}} \\ &= S(k) + \frac{k+1}{2^k(k+2)} \binom{k}{\frac{k-1}{2}} + \frac{1}{2^k(k+2)} \binom{k}{\frac{k-1}{2}} \\ &\quad - \frac{k+3}{2^{k+2}(k+4)} \binom{k+2}{\frac{k+1}{2}} \\ &= S(k) + \frac{1}{2^k} \binom{k}{\frac{k-1}{2}} - \frac{k+3}{2^{k+2}(k+4)} \binom{k+2}{\frac{k+1}{2}} \\ &\geq S(k) + \frac{1}{2^k} \binom{k}{\frac{k-1}{2}} - \frac{1}{2^{k+2}} \binom{k+2}{\frac{k+1}{2}} \\ &= S(k) + \frac{(k+1)!}{2^{k+2} \left(\frac{k+1}{2}\right)! \left(\frac{k+3}{2}\right)!} (k+3 - (k+2)) \\ &= S(k) + \frac{(k+1)!}{2^{k+2} \left(\frac{k+1}{2}\right)! \left(\frac{k+3}{2}\right)!} \end{aligned}$$

and hence $S(k+2) \geq S(k)$. This completes the proof. \square

Corollary 3.6. For $x > 0$ the series

$$\sum_{k=0}^{\infty} c_k x^k,$$

where the coefficients c_k are determined by (3.4), is alternating.

Proof. For k even we clearly have $c_k < 0$. On the other hand, for k odd c_k is positive since

$$c_k \geq \frac{\pi}{(k+2)!!}$$

by Lemma 3.5. □

The following result shows that the sum in Lemma 3.5 can also be bounded from above.

Lemma 3.7. *For all $k \geq 3$ odd, $S(k)$ defined by (3.5) satisfies $2\sqrt{2\pi}S(k) \leq \pi^2$.*

Proof. By elementary manipulations as in the proof of Lemma 3.5 we obtain

$$S(k+2) = S(k) + \frac{2\sqrt{2\pi}(k+3)}{2^{k+2}(k+2)(k+4)} \binom{k+2}{\frac{k+1}{2}}.$$

From the Stirling approximation we infer

$$\binom{k+2}{\frac{k+1}{2}} \leq \sqrt{\frac{2}{\pi}} \frac{2^{k+2}}{\sqrt{(k+2)}},$$

and hence

$$S(k+2) \leq S(k) + \frac{\sqrt{2}}{\sqrt{\pi}(k+2)^{3/2}}.$$

Moreover,

$$S(3) = 1 + \frac{1}{6} - \frac{3}{10} \leq 1 + \frac{1}{3^{3/2}}$$

allows us to get the estimates

$$\lim_{k \rightarrow \infty} 2\sqrt{2\pi}S(k) \leq 4 \sum_{k=0}^{\infty} \frac{1}{(2k+1)^{3/2}} \leq \pi^2.$$

Since S is increasing by Lemma 3.5, it follows that $S(k) \leq \pi^2$ for all $k \geq 3$ odd. This completes the proof. □

Lemma 3.8. *Let the coefficients c_k , $k = 1, 2, \dots$, be given by (3.4) and let $x > 0$ be fixed. Then the terms in the series*

$$\sum_{k=M}^{\infty} c_k x^k$$

decreases monotonically for M large enough.

Proof. Consider first the case when k is odd. Then, using the upper bound of Lemma 3.7 yields

$$\left| \frac{c_{k+1}x^{k+1}}{c_k x^k} \right| \leq \frac{\frac{\pi^2}{(k+3)!!}}{\frac{\pi^2}{(k+2)!!}} x = \frac{(k+2)!!}{(k+3)!!} x$$

which is less than one for sufficiently large k (depending on x). Similarly for k even we can use the lower bound from Lemma 3.5 to obtain

$$\left| \frac{c_{k+1}x^{k+1}}{c_kx^k} \right| \leq \frac{\frac{\pi^2}{(k+3)!!}}{\frac{\pi}{(k+2)!!}}x = \frac{(k+2)!!}{(k+3)!!}\pi x$$

which again is less than one for large enough k . This yields the claim. \square

Lemma 3.9. *Denote*

$$M(x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k+2)!!}$$

and

$$N(x) = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+3)!!}.$$

Then

$$M(x) > N(x)$$

for all $x \geq 0$.

Proof. By differentiating, we get

$$M'(x) = \sum_{k=0}^{\infty} \frac{(2k+2)x^{2k+1}}{(2k+4)!!} = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+3)!!} \frac{(k+1)}{2^{2k+2}} \binom{2k+3}{k+1}.$$

Now using

$$\binom{2k+3}{k+1} \geq \frac{2^{2k+2}}{\sqrt{k+1}}$$

gives

$$M'(x) \geq \sum_{k=0}^{\infty} \frac{\sqrt{k+1}}{(2k+3)!!} x^{2k+1} \geq N(x).$$

Similarly, it holds that

$$N'(x) = \sum_{k=0}^{\infty} \frac{(2k+1)x^{2k}}{(2k+3)!!} = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k+2)!!} \frac{(2k+1)2^{2k+2}}{(k+2)\binom{2k+3}{k+1}}.$$

In this case we can use

$$\binom{2k+3}{k+1} \geq \frac{2^{2k+3}}{2k+3}$$

leading to

$$N'(x) \leq \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k+2)!!} \frac{k+\frac{1}{2}}{(k+2)(2k+3)} \leq M(x).$$

Combining the two bounds above gives us

$$\begin{aligned} \frac{d}{dx} (M^2(x) - N^2(x)) &= 2M(x)M'(x) - 2N(x)N'(x) \\ &\geq M(x)N(x) - M(x)N(x) = 0. \end{aligned}$$

Consequently, $M^2 - N^2$ is an increasing function for $x \geq 0$, which leads to

$$M^2(x) - N^2(x) \geq M^2(0) - N^2(0) = 1.$$

It follows that

$$M(x) \geq \sqrt{1 + N^2(x)} > N(x),$$

and the proof is complete. \square

We are finally in the position to prove Proposition 2.7.

Proof of Proposition 2.7. By Corollary 3.6 and Lemma 3.8 the series expansion (3.4) for $f'(x) \exp(x^2/2)$ is convergent by the Leibniz alternating series test. Since the radius of convergence is unbounded, it is analytic, and thus we may split the series into its positive part and negative part, given by

$$\sum_{k=0}^{\infty} c_{2k} x^{2k} =: Q(x), \quad \sum_{k=0}^{\infty} c_{2k+1} x^{2k+1} =: P(x).$$

By Lemma 3.4 it holds that $Q(x) = -\pi^2 M(x)$ where $M(x)$ is as in Lemma 3.9. Now Lemma 3.7 gives

$$c_{2k} \leq \frac{\pi^2}{(k+2)!!},$$

and hence

$$Q(x) \leq \pi^2 N(x),$$

where $N(x)$ is as in Lemma 3.9. Therefore

$$f'(x) \exp(x^2/2) = Q(x) + P(x) \leq -\pi^2 M(x) + \pi^2 N(x).$$

Applying Lemma 3.9 to the above inequality now gives

$$f'(x) \exp\left(\frac{x^2}{2}\right) < 0,$$

and thus $f'(x) < 0$ for all $x \geq 0$, where f is given by (3.2). The claim follows from this. \square

A Some results from differential geometry

Here we introduce some classical concepts about manifolds and their properties used in the paper.

Let M be a smooth and compact n -dimensional manifold embedded in \mathbb{R}^k . At each point, the tangent space, $T_x M$, is equipt with the Euclidean inner product (Riemannian metric), inherited from the ambient space \mathbb{R}^k making M an isometrically embedded manifold as a *Riemannian manifold*. The Riemannian metric, denoted $\langle \cdot, \cdot \rangle$, induces a notion of length along smooth curves, $\gamma : [a, b] \rightarrow M$, by the formula

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt.$$

Given two points, or equivalently a starting point and starting velocity, the curve which minimizes this distance is called a *geodesic*. Centered at a point $x \in M$, the

map $\exp_x : T_x M \rightarrow M$ is called the Riemannian exponential map centered at $x \in M$. Given a tangent vector $v_x \in T_x M$, let γ be the unit speed geodesic such that $\gamma(0) = x$ and $\dot{\gamma}(0) = \frac{v_x}{\|v_x\|}$, then

$$\exp(v_x) = \gamma(\|v_x\|).$$

We shall assume that M is *geodesically complete*, meaning that the exponential map is defined on the whole $T_x M$ for all $x \in M$. Denote the inverse of \exp_p (restricted to all points y for which there is a unique geodesic connecting y to p) by \log_p . The distance function, $\text{dist} : M \times M \rightarrow [0, \infty)$ (which is a topological metric) is then defined as $\text{dist}(x, y) = \|\log_x(y)\|$. The Riemannian metric also gives a way to measure the volume (and orientation) of a parallelootope inside the tangent space, i.e. a rescaling of the determinant from linear algebra. This generalized determinant shall be referred to as $d\text{Vol}_M$, or the Riemannian volume form, which locally looks like the Lebesgue measure. By *vector field*, we mean a smooth assignment of a point $x \in M$ to a tangent vector $T_x M$, and the space of vector fields is denoted by $\mathfrak{X}(M)$. For a given smooth function $f \in C^\infty(M)$, a vector field X acts as a derivation on f in the following sense. Take a smooth curve $(-\varepsilon, \varepsilon) \rightarrow M$ such that $\gamma(0) = p$, $\dot{\gamma}(0) = X(p)$. Then the function $X(f) : M \rightarrow \mathbb{R}$ is point-wise defined by $X(f)(p) = (f \circ \gamma)'(0)$.

Note that for given vector fields $X, Y : M \rightarrow TM$, the expression $\langle X, Y \rangle$ is a smooth function $M \rightarrow \mathbb{R}$. In order to differentiate vector fields, a notion of *connection* is required. Here we shall use the *Levi-Civita connection* $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \rightarrow \mathfrak{X}(M)$ which is uniquely defined by the following identities:

$$\text{i) } \nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z$$

for all smooth functions $f, g \in C^\infty(M)$ and all smooth vector fields $X, Y, Z \in \mathfrak{X}(M)$;

$$\text{ii) } \nabla_X(gY + hZ) = g\nabla_X Y + X(g)Y + h\nabla_X Z + X(h)Z$$

for all smooth functions $g, h \in C^\infty(M)$ and all smooth vector fields $X, Y, Z \in \mathfrak{X}(M)$;

$$\text{iii) } X(\langle Y, Z \rangle) = \langle \nabla_X(Y), Z \rangle + \langle Y, \nabla_X(Z) \rangle$$

for all smooth vector fields $X, Y, Z \in \mathfrak{X}(M)$;

$$\text{iv) } (\nabla_X Y - \nabla_Y X)(f) = X(Y(f)) - Y(X(f))$$

as derivations for all smooth functions $f \in C^\infty(M)$ and all smooth vector fields $X, Y \in \mathfrak{X}(M)$.

Lastly, let $P_{q,\ell} : T_q M \rightarrow T_\ell M$ denote the parallel transport from q to ℓ . That is, $P_{q,\ell}(v_q)$ is the unique point $\tau(\text{dist}(q, \ell)) \in T_\ell M$ that satisfies the initial value problem

$$\nabla_{\dot{\gamma}} \tau(t) = 0, \quad \tau(0) = v_q,$$

where $\gamma : [0, \text{dist}(q, \ell)] \rightarrow M$ is the geodesic connecting q to ℓ and τ is a vector field along γ , i.e. $\tau(t) = \tau_{\gamma(t)}$. It is worthwhile to consider the parallel transport of the Riemannian logarithm map

$$q \mapsto P_{q,\ell} \log_q(v),$$

where $\nu, \ell \in M$ are fixed points. This map has the Taylor expansion around $q = \ell$ given by

$$P_{q,\ell} \log_q(\nu) = \log_\ell(\nu) + \nabla_{\log_\ell(q)} \log_\ell(\nu) + \mathcal{O}(\text{dist}(q, \ell)^2). \quad (\text{A.1})$$

Funding

This project has received partial funding from Huawei Technologies.

References

- [1] Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables vol. 55. US Government printing office (1968) [MR0415956](#)
- [2] Ahidar-Coutrix, A., Le Gouic, T., Paris, Q.: Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics. *Probab. Theory Relat. Fields* **177**(1), 323–368 (2020) [MR4095017](#). <https://doi.org/10.1007/s00440-019-00950-0>
- [3] Chang, T.: Spherical regression and the statistics of tectonic plate reconstructions. *Int. Stat. Rev.*, 299–316 (1993)
- [4] Dawson, H.G.: On the numerical value of $\int_0^h e^{x^2} dx$. *Proc. Lond. Math. Soc.* **1**(1), 519–522 (1897) [MR1576451](#). <https://doi.org/10.1112/plms/s1-29.1.519>
- [5] Fisher, N.I., Lewis, T., Embleton, B.J.J.: Statistical Analysis of Spherical Data. Cambridge University Press (1993) [MR1247695](#)
- [6] Gao, F., Chia, K., Machin, D.: On the evidence for seasonal variation in the onset of acute lymphoblastic leukemia (all). *Leuk. Res.* **31**(10), 1327–1338 (2007)
- [7] Hernandez-Stumpfhauser, D., Breidt, F.J., van der Woerd, M.J.: The general projected normal distribution of arbitrary dimension: Modeling and Bayesian inference. *Bayesian Anal.* **12**(1), 113–133 (2017) [MR3597569](#). <https://doi.org/10.1214/15-BA989>
- [8] Hotz, T., Le, H., Wood, A.T.A.: Central limit theorem for intrinsic Frechet means in smooth compact Riemannian manifolds. *Probab. Theory Relat. Fields* (2024) [MR4771114](#). <https://doi.org/10.1007/s00440-024-01291-3>
- [9] Hussien, M.T., Seddik, K.G., Gohary, R.H., Shaqfeh, M., Alnuweiri, H., Yanikomeroğlu, H.: Multi-resolution broadcasting over the Grassmann and Stiefel manifolds. In: 2014 IEEE International Symposium on Information Theory, pp. 1907–1911 (2014). IEEE
- [10] Kells, L.M., Kern, W.F., Bland, J.R.: Plane and Spherical Trigonometry. US Armed Forces Institute (1940) [MR1524230](#). <https://doi.org/10.2307/2302987>
- [11] Lang, M., Feiten, W.: Mpg-fast forward reasoning on 6 dof pose uncertainty. In: ROBOTIK 2012; 7th German Conference on Robotics, pp. 1–6 (2012). VDE
- [12] Levitt, M.: A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104**(1), 59–107 (1976)
- [13] Li, K., Frisch, D., Radtke, S., Noack, B., Hanebeck, U.D.: Wavefront orientation estimation based on progressive Bingham filtering. In: 2018 Sensor Data Fusion: Trends, Solutions, Applications (SDF), pp. 1–6 (2018). IEEE
- [14] Nuñez-Antonio, G., Ausín, M.C., Wiper, M.P.: Bayesian nonparametric models of circular variables based on Dirichlet process mixtures of normal distributions. *J. Agric. Biol. Environ. Stat.* **20**, 47–64 (2015) [MR3334466](#). <https://doi.org/10.1007/s13253-014-0193-y>

- [15] Nuñez-Antonio, G., Gutiérrez-Peña, E.: A Bayesian analysis of directional data using the projected normal distribution. *J. Appl. Stat.* **32**, 995–1001 (2005) [MR2221902](#). <https://doi.org/10.1080/02664760500164886>
- [16] Nuñez-Antonio, G., Gutiérrez-Peña, E., Escarela, G.: A Bayesian regression model for circular data based on the projected normal distribution. *Stat. Model.* **11**, 185–201 (2011) [MR2849683](#). <https://doi.org/10.1177/1471082X1001100301>
- [17] Pennec, X.: Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.* **25**, 127–154 (2006) [MR2254442](#). <https://doi.org/10.1007/s10851-006-6228-4>
- [18] Pennec, X., Arsigny, V.: Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups. In: *Matrix Information Geometry*, pp. 123–166. Springer (2012) [MR2964451](#). https://doi.org/10.1007/978-3-642-30232-9_7
- [19] Pitaval, R., Tirkkonen, O.: Joint Grassmann-Stiefel quantization for MIMO product codebooks. *IEEE Trans. Wirel. Commun.* **13**(1), 210–222 (2013)
- [20] Presnell, B., Morrison, S.P., Littell, R.C.: Projected multivariate linear models for directional data. *J. Am. Stat. Assoc.* **93**, 1068–1077 (1998) [MR1649201](#). <https://doi.org/10.2307/2669850>
- [21] Seddik, K.G., Gohary, R.H., Hussien, M.T., Shaqfeh, M., Alnuweiri, H., Yanikomeroglu, H.: Multi-resolution multicasting over the Grassmann and Stiefel manifolds. *IEEE Trans. Wirel. Commun.* **16**(8), 5296–5310 (2017)
- [22] Sveier, A., Egeland, O.: Pose estimation using dual quaternions and moving horizon estimation. *IFAC-PapersOnLine* **51**(13), 186–191 (2018)
- [23] Tanabe, A., Fukumizu, K., Oba, S., Takenouchi, T., Ishii, S.: Parameter estimation for von mises–fisher distributions. *Comput. Stat.* **22**, 145–157 (2007) [MR2299252](#). <https://doi.org/10.1007/s00180-007-0030-7>
- [24] Wang, F., Gelfand, A.E.: Directional data analysis under the general projected normal distribution. *Stat. Methodol.* **10**(1), 113–127 (2013) [MR2974815](#). <https://doi.org/10.1016/j.stamet.2012.07.005>