Isotropic Covariance Functions on Spheres: Some Properties and Modeling Considerations

Joseph Guinness[a], Montserrat Fuentes[a]

[a] *NC State University, Department of Statistics*

Abstract

Introducing flexible covariance functions is critical for interpolating spatial data since the properties of interpolated surfaces depend on the covariance function used for Kriging. An extensive literature is devoted to covariance functions on Euclidean spaces, where the Matérn covariance family is a valid and flexible parametric family capable of controlling the smoothness of corresponding stochastic processes. Many applications in environmental statistics involve data located on spheres, where less is known about properties of covariance functions, and where the Matérn is not generally a valid model with great circle distance metric. In this paper, we advance the understanding of covariance functions on spheres by defining the notion of and proving a characterization theorem for m times mean square differentiable processes on d-dimensional spheres. Stochastic processes on spheres are commonly constructed by restricting processes on Euclidean spaces to spheres of lower dimension. We prove that the resulting sphere-restricted process retains its differentiability properties, which has the important implication that the Matérn family retains its full range of smoothness when applied to spheres so long as Euclidean distance is used. The restriction operation has been questioned for using Euclidean instead of great circle distance. To address this question, we construct several new covariance functions and compare them to the Matérn with Euclidean distance on the task of interpolating smooth and non-smooth datasets. The Matérn with Euclidean distance is not outperformed by the new covariance functions or the existing covariance functions, so we recommend using the Matérn with Euclidean distance due to the ease with which it can be computed.

Keywords: Kriging, Fourier series, positive definite functions

1. Introduction

When modeling dependent spatial data, the covariance function used is crucial for producing accurate predictions and estimating prediction uncertainties.

Email address: jsguinne@ncsu.edu (Joseph Guinness)

Statistical theory (Stein, 1999; Zhang, 2004) shows that when the goal is interpolation of highly dependent data packed tightly on a compact domain, it is of utmost importance to correctly specify the local properties of the process, which are determined by the behavior of the covariance function near the origin. In recent years the Matérn family of covariance functions has gained widespread popularity in spatial statistics (Guttorp and Gneiting, 2006) due partly to its ability to control the local behavior of the process. Specifically, let $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^d$, be a random field. The isotropic Matérn covariance function is given by

$$M(\|\boldsymbol{h}\|) = \mathrm{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{h})) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}\mathcal{K}_\nu(\alpha\|\boldsymbol{h}\|)(\alpha\|\boldsymbol{h}\|)^\nu, \qquad (1)$$

where $\sigma^2, \alpha, \nu > 0$, and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind. The popularity of the Matérn is also due partly to this representation, which allows the function to be computed by way of rapidly converging series expansions for the Bessel function (Digital Library of Mathematical Functions, 2012, Chapter 10). We say that M is isotropic because it depends on the locations $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{h}$ only through the Euclidean distance $\|\boldsymbol{h}\|$ between them. The parameter of interest here is ν, which controls the smoothness of the process, defined in terms of its mean square differentiability: a process on a Euclidean space that has covariance function M has m mean square derivatives if and only if $\nu > m$.

In environmental statistics, we often encounter data associated with locations on the surface of the Earth, for example observations from satellites or output from climate models, and in astronomy and cosmology, the observations are often associated with an azimuth and altitude in the sky, so it is important to introduce flexible classes of covariance functions that are valid on spheres. Marinucci and Peccati (2011) provide a broad overview of the theory of random fields on spheres. The question of validity of covariance functions on spheres has been studied extensively by Huang et al. (2011) and further by Gneiting (2013), who proved that many of the commonly used covariance functions on Euclidean spaces are valid on spheres when Euclidean distance is replaced by great circle distance–the more natural distance metric on a sphere. However, the Matérn is positive definite with great circle distance only if $\nu \leq 1/2$. The fact that the validity of the Matérn on spheres is tied to the value of the smoothness parameter handcuffs its usefulness for modeling a wide range of smooth and non-smooth spatial data. Recently there have been efforts to introduce new covariance functions on spheres. Ma (2012), Du et al. (2013), and Ma (2014) provide closed-form covariance functions and variogram functions for vector-valued processes on spheres, and Heaton et al. (2014) defines covariance functions on spheres in terms of kernel convolutions. However, none of the functions any of these authors studied possess the flexibility to specify the smoothness of the process like the Matérn does. An exception is Jeong and Jun (2015), who introduce a "Matérn-like" covariance function on spheres but this covariance function must be approximated and does not outperform much simpler alternatives.

Spheres are subsets of Euclidean spaces, so a covariance function that is valid on a Euclidean space–such as the Matérn–can be applied to a sphere of lower dimension if the Euclidean distance is used. More formally, for $d \geq 1$, define the d-sphere as $\mathbb{S}^d = \{\boldsymbol{x} \in \mathbb{R}^{d+1} : \|\boldsymbol{x}\| = 1\}$ and the great circle distance metric as $\theta(\boldsymbol{x}, \boldsymbol{y}) = \arccos(\langle \boldsymbol{x}, \boldsymbol{y} \rangle)$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^d$, where $\langle \cdot, \cdot \rangle$ denotes the usual inner product on $\mathbb{R}^d$. The Euclidean distance between two points on a sphere, which is also known as the chordal distance, can be expressed in terms of great circle distance as $\|\boldsymbol{x} - \boldsymbol{y}\| = 2\sin(\theta(\boldsymbol{x}, \boldsymbol{y})/2)$, so if K is a valid isotropic covariance function on $\mathbb{R}^{d+1}$, then $\psi(\theta) = K(2\sin(\theta/2))$ is a valid covariance function on $\mathbb{S}^d$ (Yadrenko, 1983; Yaglom, 1987). Stated more simply, this approach starts with a valid process on $\mathbb{R}^{d+1}$ and restricts it to the sphere $\mathbb{S}^d$, so while the process on $\mathbb{S}^d$ is trivially valid, we must use the chordal distance in calculations of the covariance. In what follows, we refer to the Matérn with chordal distance, $\varphi(\theta) = M(2\sin(\theta/2))$, as the chordal Matérn covariance function.

Our work is concerned with understanding the properties of covariance functions on spheres, specifically with respect to mean square differentiability, and exploring the modeling capabilities of the chordal Matérn with real data. In Section 2, we define the notion of a mean square differentiable process on a sphere and provide a concise theorem characterizing m times mean square differentiable processes in terms of their covariance functions and Fourier series. Since it is common to use the restriction construction to define valid covariance functions on spheres, we prove a corollary stating that the process restricted to a sphere retains the differentiability properties of the original process. This result has the important implication that the chordal Matérn retains the full flexibility that the Matérn does, in terms of smoothness.

While the restriction operation is convenient due to the abundance of flexible models on Euclidean spaces, such as the Matérn, Gneiting (2013) argued that this "may result in physically unrealistic distortions." It is important to understand the implications of defining a covariance function in terms of chordal distance and whether such covariance functions will poorly model data observed over large regions on a sphere. There have been some efforts to compare the two distance metrics, most notably Banerjee (2005), who fits parametric spatial covariance functions to data observed at locations on the Earth. The results there suggested that using the chordal versus great circle distance may produce slightly different model estimates. However, the observation region for those data was quite small compared to the entire globe–less than $5^\circ \times 5^\circ$ latitude $\times$ longitude–and the study considered the Matérn with great circle distance, which is not generally a positive definite function on a sphere, so a more thorough investigation is warranted.

To provide insight into the appropriateness of the chordal Matérn for data observed on spheres, we compare its ability to model and interpolate smooth and non-smooth datasets to that of a number of existing and new covariance functions that we introduce in Section 3. The existing covariance functions were introduced in Huang et al. (2011) and Gneiting (2013) and consist of covariance functions that are valid on Euclidean spaces that remain valid on spheres when Euclidean distance is replaced by great circle distance. In Section 3, we introduce several new families of covariance functions capable of modeling smooth and nonsmooth processes and whose constructions respect circular and spherical geometry. Using existing theoretical results and our new theoretical results we outline the differentiability properties of the new covariance functions. We also discuss computational considerations and provide closed forms for the covariance functions in some cases, one of which corresponds to the characteristic function of an integer-valued version of the t-distribution.

The covariance functions are compared in Section 4 on two datasets of different degrees of smoothness, both of which span large distances around the Earth. We find that for the problem of dense interpolation, the chordal Matérn is not outperformed by any of the new or existing covariance functions, and it is sometimes a substantial improvement in terms of loglikelihood and predictive performance over existing covariance functions that take great circle distance as the argument. Finally, we conclude in Section 5 with a discussion of our work and practical practical recommendations for choosing covariance functions to model data observed on spheres.

2. Mean Square Differentiability on Spheres

Following Gneiting (2013), a function $h : \mathbb{S}^d \times \mathbb{S}^d \to \mathbb{R}$ is positive definite on the d-sphere if

$$\sum_{j=1}^{n}\sum_{k=1}^{n} c_j c_k h(\boldsymbol{x}_j, \boldsymbol{x}_k) \geq 0 \tag{2}$$

for all n, locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{S}^d$, and constants $c_1, \ldots, c_n \in \mathbb{R}$. A function is strictly positive definite if the inequality in (2) is strict whenever at least one of the constants is non-zero. Positive definiteness ensures that the variance of all linear combinations of observations is positive. The function h is isotropic if there exists a function $\psi : [0, \pi] \to \mathbb{R}$ such that

$$h(\boldsymbol{x}, \boldsymbol{y}) = \psi(\theta(\boldsymbol{x}, \boldsymbol{y})) \quad \text{for all} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{S}^d.$$

It is important to note that the assumption of isotropy is often not justifiable for data on the surface of the Earth. However, many existing methods for generating anisotropic covariance functions arise from making modifications to isotropic covariance functions. This includes deformations (Sampson and Guttorp, 1992; Anderes and Stein, 2008), partial differential equation approaches (Jun and Stein, 2007), and convolutions (Higdon, 1998; Paciorek and Schervish, 2006). Thus, even if one is ultimately interested in anisotropic models, careful study of the properties of isotropic models remains vital.

As discussed in the introduction, correctly specifying the local properties of a process is important when one is interested in interpolating spatial data and providing accurate estimates of prediction uncertainty. Marinucci and Peccati (2013) proved that isotropy of a random field on a compact group (of which spheres are important examples) entails mean square continuity of the random field. The local properties of a process can be further described with respect to the number of mean square derivatives it possesses. For isotropic processes on

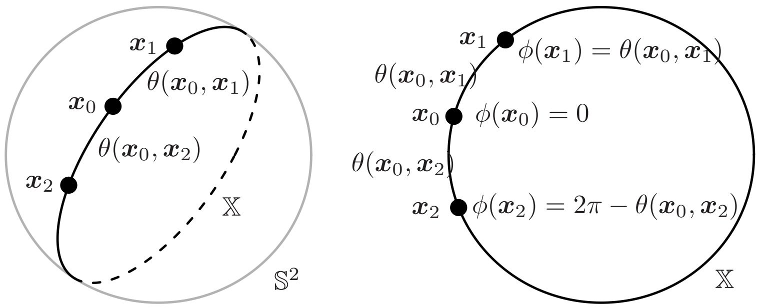

Figure 1: Illustration of a great circle $\mathbb{X}$ on the two-sphere $\mathbb{S}^2$ and the definition of ϕ.

$\mathbb{S}^2$, Hitczenko and Stein (2012) gave conditions for mean square differentiability when the covariance function is expressed in terms of its spherical harmonic representation. Lang and Schwab (2013) give further results on Hölder continuity and differentiability of sample paths of the process. Here, we give conditions that can be used to determine the number of mean square derivatives directly from the covariance function or from its Fourier series representation, and the results apply to spheres of arbitrary dimension.

Let $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{S}^d$ be a stochastic process on the d-sphere with isotropic covariance function ψ, and let $\mathcal{H}_\psi$ be the Hilbert space of linear combinations of Z with finite variance. Thus $\mathcal{H}_\psi$ is the set of all random variables with finite variance that can be expressed as $\sum_{k=1}^{n} a_k Z(\boldsymbol{x}_k)$ with $a_k \in \mathbb{R}$ and n possibly infinite. Derivatives of random or deterministic functions on Euclidean spaces must be defined with respect to a direction along a straight line. On spheres, the analog of a straight line is a geodesic or a great circle, so to study the derivatives of random functions on spheres, we must define the notion of a great circle, which is the intersection of $\mathbb{S}^d$ with any plane that passes through the origin. A sphere contains infinitely many great circles. Examples can be described using the analogy of the Earth as a sphere; lines of longitude and the equator lie along great circles, whereas non-equatorial lines of latitude do not.

Suppose $\mathbb{X}$ is the collection of all the points along one great circle. The great circle $\mathbb{X}$ and $\mathbb{S}^1$ are isometric since rotations in Euclidean spaces are isometric isomorphisms, and $\mathbb{S}^1$ and $[0, 2\pi)$ are isometric if we define distance in $[0, 2\pi)$ to be $d(\phi_1, \phi_2) = \min(|\phi_1 - \phi_2|, 2\pi - |\phi_1 - \phi_2|)$. Thus there there is a distance-preserving mapping $\phi : \mathbb{X} \to [0, 2\pi)$ that associates each point on a great circle with a unique angle. In Figure 1, we show an example with $\phi(\boldsymbol{x}_0) = 0$ and a "clockwise" orientation.

Next, for some choice of ϕ, we define $Z_{\mathbb{X}}(\phi(\boldsymbol{x})) = Z(\boldsymbol{x})$ to be the restriction of Z to $\mathbb{X}$. Then we say that $Z_{\mathbb{X}}$ is mean square differentiable at $\boldsymbol{x}$ if the limit

$$Z_{\mathbb{X}}^{(1)}(\phi(\boldsymbol{x})) = \lim_{\varepsilon \to 0} \frac{Z_{\mathbb{X}}(\phi(\boldsymbol{x}) + \varepsilon) - Z_{\mathbb{X}}(\phi(\boldsymbol{x}))}{\varepsilon}$$

exists in $\mathcal{H}_{\psi}$, and we say that Z is mean square differentiable at $\boldsymbol{x}$ if $Z_{\mathbb{X}}^{(1)}(\phi(\boldsymbol{x}))$ exists for every $\mathbb{X}$ that contains $\boldsymbol{x}$. The entire process Z is mean square differentiable if Z is mean square differentiable at $\boldsymbol{x}$ for every $\boldsymbol{x} \subset \mathbb{S}^d$. Clearly, if the process is isotropic, mean square differentiability at one point along one great circle implies mean square differentiability of the entire process. To define higher order differentiability, we say that Z is $m+1$ times mean square differentiable at $\boldsymbol{x}$ if Z is m times mean square differentiable, and

$$Z_{\mathbb{X}}^{(m+1)}(\phi(\boldsymbol{x})) = \lim_{\varepsilon \to 0} \frac{Z_{\mathbb{X}}^{(m)}(\phi(\boldsymbol{x}) + \varepsilon) - Z_{\mathbb{X}}^{(m)}(\phi(\boldsymbol{x}))}{\varepsilon}$$

exists for every $\mathbb{X}$ that contains $\boldsymbol{x}$. Then Z is $m+1$ times mean square differentiable if Z is $m+1$ times mean square differentiable at $\boldsymbol{x}$ for every $\boldsymbol{x} \subset \mathbb{S}^d$.

Armed with a proper definition of mean square differentiability on spheres, it is straightforward to prove the following theorem characterizing isotropic mean square differentiable processes on spheres.

Theorem 1. *If an isotropic process Z on $\mathbb{S}^d$ has covariance function*

$$\psi(\theta) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} f_k \exp(ik\theta),$$

then the following statements are equivalent:

1. *Z is m times mean square differentiable.*
2. $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$.
3. *$\psi^{2m}(0)$ exists and is finite.*

The proof is given in Appendix A. Condition 2 on the Fourier coefficients is useful both in practice, since in Section 3 we show that valid covariance functions on $\mathbb{S}^3$ are easily specified via their Fourier coefficients, and in theory since every covariance function is embeddable in a 2π-periodic function and thus has a Fourier series representation.

It is common practice to restrict a process on a Euclidean space to a sphere of lower dimension, so it is of interest to know how this operation affects the mean square differentiability properties of the process on a sphere. The following corollary follows from Theorem 1 and gives credence to the idea that restricting a process to a sphere does not distort the local properties of the process.

Corollary 1. *If $K(h)$ is the covariance function for an isotropic, m times mean square differentiable process on $\mathbb{R}^{d+1}$, then $\psi(\theta) = K(2\sin(\theta/2))$ is the covariance function for an isotropic, m times mean square differentiable process on $\mathbb{S}^d$.*

Proof. Since K is the covariance function for an m times mean square differentiable process on $\mathbb{R}^{d+1}$, the derivatives $K^{(j)}(0)$ exist and are finite for all $j \leq 2m$. The function $\psi(\theta) = K(2\sin(\theta/2))$ is always the covariance function for an isotropic process on $\mathbb{S}^d$. According to Theorem 1, to prove that the resulting process is m times mean square differentiable on $\mathbb{S}^d$, we must show that $\psi^{(2m)}(0)$ exists and is finite. Writing $f(\theta) = 2\sin(\theta/2)$, by Faà di Bruno's formula for derivatives of composite functions (Johnson, 2002),

$$\psi^{(2m)}(\theta) = \sum \frac{(2m)!}{b_1!b_2!\cdots b_{2m}!} K^{(j)}(f(\theta)) \left(\frac{f'(\theta)}{1!}\right)^{b_1} \left(\frac{f''(\theta)}{2!}\right)^{b_2} \cdots \left(\frac{f^{(2m)}(\theta)}{(2m)!}\right)^{b_{2m}},$$

where the sum is over all nonnegative integer solutions $b_1, \ldots, b_{2m}$ of $b_1 + 2b_2 + \cdots + 2mb_{2m} = 2m$, with $j = b_1 + \cdots + b_{2m}$. The largest value of j that appears in the sum is when $b_1 = 2m$ and $b_i = 0$ for all $i \neq 1$, which gives $j = 2m$. Since $K^{(j)}(0)$ exists and is finite for all $j \leq 2m$, $\psi^{(2m)}(0)$ must exist and be finite as well, since powers of the derivatives of $f(\theta) = 2\sin(\theta/2)$ are analytic, and the sum has finitely many terms. □

Corollary 1 shows that one can construct valid covariance functions on spheres with varying degrees of smoothness by restricting flexible covariance functions that are valid on Euclidean spaces, like the Matérn for example, to spheres of lower dimension. As noted in the introduction, some authors have questioned this practice since it requires the use of the chordal distance instead of the more natural great circle distance. To address this issue, we construct in Section 3 several new covariance functions whose construction respects circular and spherical geometry, and in Section 4 we compare the various models to each other and to a number of existing covariance functions on two sets of data.

3. Construction of Flexible Covariance Functions on Spheres

Commonly, practitioners analyze data associated with locations on the surface of the Earth, so it is of interest to study covariance functions that are valid on $\mathbb{S}^2$ specifically. The usual method of specifying an isotropic function on $\mathbb{S}^2$ is through the coefficients $\{b_k\}$ in its Legendre polynomial representation,

$$\psi(\theta) = \sum_{k=0}^{\infty} b_k P_k(\cos\theta), \tag{3}$$

where P_k is the kth Legendre polynomial. Schoenberg (1942) proved that b_k nonnegative and summable guarantees that ψ is nonnegative definite on $\mathbb{S}^2$. Terdik (2013) gives numerous examples of covariance functions constructed with the representation in (3). Using Theorem A in Hitczenko and Stein (2012), it is straightforward to write down flexible classes of covariance functions on $\mathbb{S}^2$, since the number of mean square derivatives of a process with covariance function (3) is controlled by the rate of decay of $\{b_k\}$. Consider the following example:

Example 1:

$$\psi(\theta) = \sum_{k=0}^{\infty} \frac{\sigma^2}{(\alpha^2 + k^2)^{\nu+1/2}} P_k(\cos\theta) \tag{4}$$

This covariance function has three parameters $\sigma^2, \alpha, \nu > 0$, with σ^2 controlling the variance, $1/\alpha$ controlling the spatial range, and ν controlling the smoothness; Theorem A in Hitczenko and Stein (2012) can be used to show that processes with this covariance function are m times mean square differentiable if and only if $\nu > m$. In this paper, we refer to the covariance function in (4) as the *Legendre-Matérn* covariance function. The connection to the Matérn covariance function should be clear to those familiar with its spectral density, which is proportional to $(\alpha^2 + \omega^2)^{-\nu-1/2}$. The practical use of Legendre polynomial representations is limited due to the difficulty of obtaining closed-form expressions for the infinite sum. We consider truncations of the sum in (4) and write Legendre-Matérn (N) to denote a truncation after N terms.

Truncations of (3) allow for implementation of the covariance functions, but truncations are not strictly positive definite since only a finite number of b_k are nonzero (Chen et al., 2003), which could lead to exactly singular covariance matrices. Further, if ν is small in (4), the sum is slow to converge, and thus a prohibitively large number of terms could be required to achieve a desired accuracy. We present a concrete example of this problem in Section 4. For these reasons, we consider a Fourier series representation due to the abundance of analytic results on Fourier series:

$$\psi(\theta) = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} f_k \exp(ik\theta). \tag{5}$$

Since $\psi(\theta)$ must be real and even, we require f_k real and even as well. Gneiting (2013) showed that if in addition, the coefficients f_k are positive, summable, and $f_k - f_{k+2} > 0$ for every $k \geq 0$, then $\psi(\theta)$ is positive definite on $\mathbb{S}^3$, and therefore $\mathbb{S}^2$ as well since $\mathbb{S}^2 \subset \mathbb{S}^3$. Stronger monotonicity conditions imply positive definiteness on higher-order spheres, and we refer the reader to Gneiting (2013) for details.

The following example shows that covariance functions that are positive definite on $\mathbb{S}^3$ can be constructed from the Bernoulli polynomials of even integer order $B_{2n}(x)$:

Example 2:

$$\psi(\theta) = \sigma^2 \left[(1+\alpha) + \frac{(2\pi)^{2n} B_{2n}(\theta/(2\pi))}{(-1)^{n-1}(2n)!} \right] = \sigma^2 \left[(1+\alpha) + \sum_{k\neq 0} \frac{\exp(ik\theta)}{|k|^{2n}} \right] \tag{6}$$

This covariance function has parameters $\sigma^2, \alpha > 0$, and $n \in \mathbb{N}$. Here, σ^2 controls the variance, and α provides an overall vertical shift to the covariance functions. The covariance function remains valid when n takes on arbitrary values greater than 1/2, but the closed form in (6) requires $n \in \mathbb{N}$. The Fourier coefficients are monotonic and have $2m$ moments if and only if $n \geq m+1$, and thus the covariance function is positive definite on $\mathbb{S}^3$ and corresponds to processes with m mean square derivatives if and only if $n \geq m+1$. We refer to the covariance function in (6) as the *Bernoulli* covariance function. It is easily computable since the Bernoulli polynomials are available in closed form.

We propose a final covariance function that we consider to be of practical use for making comparisons with the chordal Matérn.

Example 3:

$$\psi(\theta) = \frac{\sigma^2}{2\pi} \sum_{k \in \mathbb{Z}} \frac{\exp(ik\theta)}{(\alpha^2 + k^2)^{\nu+1/2}}, \tag{7}$$

The three parameters are $\sigma^2, \alpha, \nu > 0$, with the same interpretations as in the Matérn and Legendre-Matérn covariance functions. In this paper, we refer to the covariance function in (7) as the *circular Matérn* covariance function. The circular Matérn is positive definite on $\mathbb{S}^1$ since $f_k = (\alpha^2 + k^2)^{-\nu-1/2}$ is positive and summable, and further, it is strictly positive definite on $\mathbb{S}^2$ and $\mathbb{S}^3$ since f_k is strictly monotone decreasing for $k \geq 0$. Using Theorem 1, processes with the circular Matérn covariance function are m times mean square differentiable if and only if $\nu > m$.

When the smoothness parameter ν is a half-integer, the circular Matérn is given by the formula

$$\psi_n(\theta) = \frac{a}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\exp(ik\theta)}{(\alpha^2 + k^2)^n},$$

where for simplicity, we set $\sigma^2 = a = 2\alpha \sinh(\alpha\pi)$, which does not affect the generality of the following result that establishes a closed form for $\psi_n(\theta)$ in terms of polynomials and hyperbolic functions.

Theorem 2.

$$\psi_n(\theta) = \sum_{k=0}^{n-1} a_{nk} (\alpha(\theta - \pi))^k \text{hyp}^{(k)}(\alpha(\theta - \pi)),$$

where $\text{hyp}^{(k)}(t)$ *is* $\cosh(t)$ *when* k *is even and* $\sinh(t)$ *when* k *is odd.*

For example,

$$\begin{aligned}
\psi_1(\theta) =& a_{10}\cosh(\alpha(\theta-\pi)),\\
\psi_2(\theta) =& a_{20}\cosh(\alpha(\theta-\pi)) + a_{21}(\alpha(\theta-\pi))\sinh(\alpha(\theta-\pi)),\\
\psi_3(\theta) =& a_{30}\cosh(\alpha(\theta-\pi)) + a_{31}(\alpha(\theta-\pi))\sinh(\alpha(\theta-\pi))+\\
& a_{32}(\alpha(\theta-\pi))^2\cosh(\alpha(\theta-\pi)),\\
\psi_4(\theta) =& \dots
\end{aligned}$$

The proof of this theorem is lengthy, but interesting, so we include it in full in Appendix S1, which also contains explicit formulas for the coefficients a_{nk}. Given the form of the Fourier coefficients of the circular Matérn, $\psi_n(\theta)/\psi_n(0)$ is the characteristic function for an integer-valued version of the t-distribution with integer degrees of freedom. The existence of a closed form in terms of polynomials and hyperbolic functions is analogous to the Matérn as well, since it has a closed form expression in terms of polynomials and exponential functions when ν is a half-integer. The details of the derivation of the closed form are provided in Appendix S1, as well as explicit formulas for the coefficients a_{nk}. We also provide computationally efficient and theoretically grounded methods for approximating the circular Matérn for arbitrary ν in Appendix S2.

4. Application to satellite and climate model data

Both Gneiting (2013) and Huang et al. (2011) list several covariance functions that are known to be valid on $\mathbb{S}^2$. In Table 1, we give the functional form and parameters for those covariance functions and for the new covariance functions discussed herein. We investigate the performance of these covariance functions on two datasets, where performance is judged based on two criteria: (1) the value of the Gaussian loglikelihood function at its maximum when fit to each dataset, and (2) the width and coverage of 90% prediction intervals for the process at held-out spatial locations. The first criterion assesses whether the covariance model gives an accurate representation of the process, and the second evaluates the accuracy of predictions and prediction uncertainty.

The first dataset we consider contains values of total column ozone derived from observations made by the Ozone Monitoring Instrument on board NASA's Aura Satellite. Aura follows a nearly sun-synchronous orbit with a period of roughly 100 minutes. We consider observations from a single orbit that encompass a wide range of latitudes over a longitude band of roughly 23 degrees near the equator. The data are plotted against latitude in Figure 2. Since all of the data are collected within a 50-minute window, we ignore the time dimension in the data and proceed as if they were collected simultaneously. We fit the covariance models to 1000 of these ozone values and for prediction purposes hold out an additional 1000 values at distinct locations. All of the data are publicly available on the web; we downloaded them using the Simple Subset Wizard (`http://disc.sci.gsfc.nasa.gov/SSW/`) with keyword OMDOAO3 for the date of March 19, 2012.

Name	Expression	Parameter Values
Chordal Matérn	$\left(\alpha 2\sin\left(\frac{\theta}{2}\right)\right)^{\nu}\mathcal{K}_{\nu}\left(\alpha 2\sin\left(\frac{\theta}{2}\right)\right)$	$\alpha, \nu > 0$
Circular Matérn	$\sum_{k=-\infty}^{\infty}(\alpha^2+k^2)^{-\nu-1/2}\exp(ik\theta)$	$\alpha, \nu > 0$
Legendre-Matérn (N)	$\sum_{k=0}^{N}(\alpha^2+k^2)^{-\nu-1/2}P_k(\cos\theta)$	$\alpha, \nu > 0$
Bernoulli	$(1+\alpha)+\sum_{k\neq 0}\lvert k\rvert^{-2n}\exp(ik\theta)$	$\alpha > 0, n \in \mathbb{N}$
Powered Exponential	$\exp\left(-(\alpha\theta)^{\nu}\right)$	$\alpha > 0; \nu \in (0,1]$
Generalized Cauchy	$\left(1+(\alpha\theta)^{\nu}\right)^{-\tau/\nu}$	$\alpha, \tau > 0; \nu \in (0,1]$
Multiquadric	$(1-\tau)^{2\alpha}/(1+\tau^2-2\tau\cos(\theta))^{\alpha}$	$\alpha > 0; \tau \in (0,1)$
Sine Power	$1-\left(\sin\frac{\theta}{2}\right)^{\nu}$	$\nu \in (0,2)$
Spherical	$\left(1+\frac{\alpha\theta}{2}\right)(1-\alpha\theta)_+^2$	$\alpha > 0$
Askey	$(1-\alpha\theta)_+^{\tau}$	$\alpha > 0; \tau \geq 2$
C^2-Wendland	$(1+\tau\alpha\theta)(1-\alpha\theta)_+^{\tau}$	$\alpha \geq \frac{1}{\pi}; \tau \geq 4$
C^4 Wendland	$\left(1+\tau\alpha\theta+\frac{\tau^2-1}{3}(\alpha\theta)^2\right)(1-\alpha\theta)_+^{\tau}$	$\alpha \geq \frac{1}{\pi}; \tau \geq 6$

Table 1: List of covariance functions. Those below the horizontal line appear in Huang et al. (2011) or Gneiting (2013)

The second dataset contains 10 meter height surface temperature outputs from a single run of the Community Climate System Model Version 4 (CCSM4). We consider a spatial field consisting of a single year's average temperature, corresponding to year 50 of this particular run of the model. The values are plotted against latitude in Figure 3. The temperature values from the climate model output tend to be smoother as a function of spatial location than are the total column ozone values. Again, we consider 1000 temperature values and hold out an additional 1000 values. The locations of the observed and held-out values are regularly-spaced over the oceans.

For each covariance function $\varphi(\theta)$ listed in Table 1, we fit the covariance function $\gamma\mathbf{1}_{\theta=0}+\sigma^2\varphi(\theta)$, so in addition to the parameters listed in Table 1, we include the possibility of a nugget term and a multiplicative term. The models are fit using residual maximum likelihood (REML), assuming an unknown mean function that is cubic in latitude. In Table 2, we report the maximum residual loglikelihood for each dataset and for each covariance family, relative to that of the best-fitting covariance function.

Conditional on the fitted covariance functions for each covariance family and each dataset, we compute best linear unbiased predictions (BLUPs) $\widehat{Z}(\boldsymbol{x}_0)$ of the process $Z(\boldsymbol{x}_0)$ at each held-out location, $\boldsymbol{x}_0$. For each prediction, we compute the mean square prediction error $E(\widehat{Z}(\boldsymbol{x}_0)-Z(\boldsymbol{x}_0))^2$ and form 90% prediction intervals based on a Gaussian assumption. The details of the computation of the BLUPs and their mean square errors can be found in Stein (1999, Sec. 1.5).

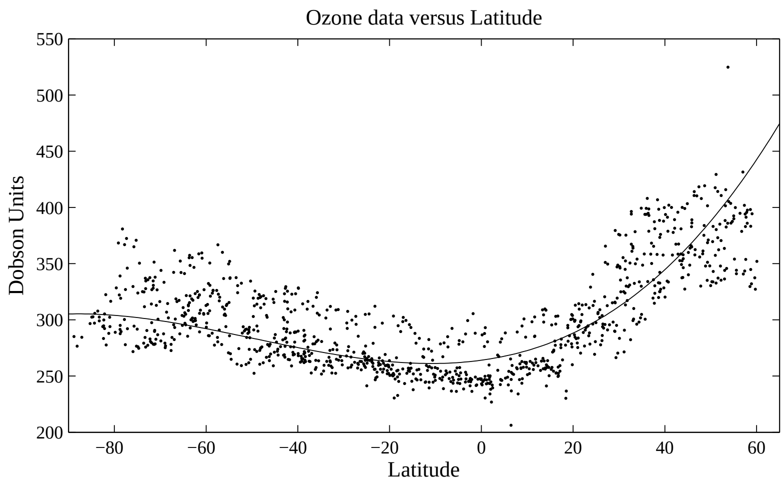

Figure 2: Total column ozone observations as a function of latitude. The smooth curve is a generalized least squares estimate of a cubic mean function, assuming the covariance function is the REML estimate of the circular Matérn.

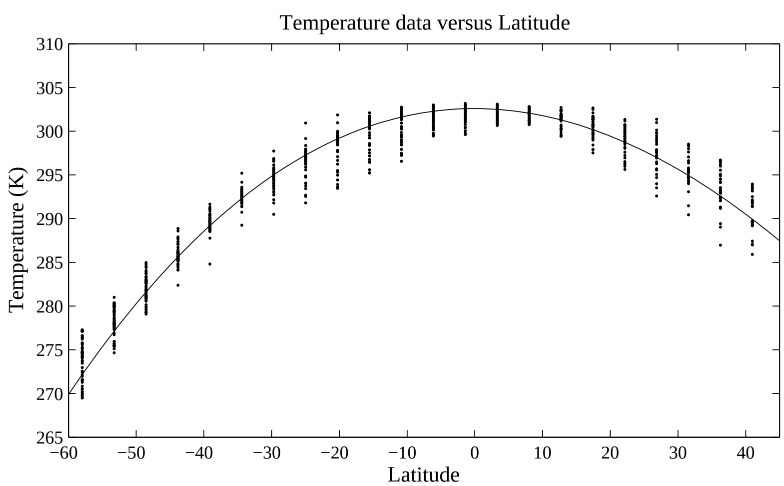

Figure 3: Climate model temperature values as a function of latitude. The smooth curve is a generalized least squares estimate of a cubic mean function, assuming the covariance function is the REML estimate of the circular Matérn.

	Ozone Data			Climate Model Data		
Cov. Family	Δloglik	width	coverage	Δloglik	width	coverage
Chordal Matérn	−0.79	42.24	90.90	−0.05	0.90	89.30
Circular Matérn	−0.79	42.24	90.90	0.00	0.90	89.30
Legendre-Matérn(100)	−14.33	43.42	91.30	−0.55	0.90	89.30
Legendre-Matérn(1000)	0.00	42.20	91.10	−0.54	0.90	89.30
Bernoulli, n=1	−6.81	41.99	90.70	−117.44	1.47	94.40
Bernoulli, n=2	−103.96	44.23	90.60	−29.86	0.96	90.00
Powered Exponential	−0.81	42.24	90.90	−119.27	1.48	94.40
Generalized Cauchy	−0.84	42.24	91.00	−119.83	1.48	94.40
Multiquadric	−9.72	43.14	91.10	−4.46	0.95	90.10
Sine Power	−2.03	42.22	91.20	−17.53	0.98	89.80
Spherical	−2.93	41.82	90.80	−121.84	1.48	94.70
Askey	−2.37	41.99	90.80	−113.66	1.47	94.40
C^2 Wendland	−16.05	43.12	91.20	−0.10	0.91	89.30
C^4 Wendland	−21.83	43.37	91.20	−8.41	0.99	90.00

Table 2: Residual maximum loglikelihoods and average widths and empirical coverage probabilities of 90% prediction intervals over all held-out sites $\boldsymbol{x}_0$. The loglikelihood is reported relative to the best-fitting model for each dataset.

In Table 2, we report the average width of the prediction intervals and their empirical coverage for each covariance function.

We see that for the ozone data, which is rougher, many of the covariance families perform nearly equally well in terms of residual maximum likelihood and prediction. The circular and chordal Matérn both return estimates of $\widehat{\nu} = 0.419$ rounded to three decimals. Thus it is not surprising that the covariance families that perform well can all be made linear at the origin. On the contrary, those that perform poorly on the ozone data–the Bernoulli with $n = 2$, the Multiquadric, the C^2-Wendland, and the C^4-Wendland–are always twice differentiable at the origin. For those three covariance functions, either the prediction interval width or the empirical coverage probabilities are too large. The Legendre-Matérn (1000) performs well, but the Legendre-Matérn (100) is suboptimal in terms of loglikelihood and prediction, suggesting that truncation after 100 terms is not sufficient in this case.

Except for the chordal Matérn, the circular Matérn, and the Legendre-Matérn, all of the covariance families that performed well on the ozone data perform poorly on the climate model data, with some performing extremely poorly. The chordal Matérn, the circular Matérn, and the Legendre-Matérn return estimates $\widehat{\nu} = 1.457$, $\widehat{\nu} = 1.459$, and $\widehat{\nu} = 1.4603$, respectively, which suggests that the climate model data can be modeled by a process that is once but not twice mean square differentiable. The Legendre-Matérn (100) and (1000) have nearly equal performance, which suggests that truncation after 100 terms is nearly sufficient in this case, which is not surprising because the coefficients for the fitted models decay more quickly with k than they did for the ozone data model. For the climate model data, the covariance families that can be made differentiable at the origin perform well, while those that are always linear at the origin perform particularly poorly in terms of loglikelihood and prediction. These covariance functions overestimate the mean square prediction error, leading to overly conservative prediction intervals. It is also interesting to note that the C^2-Wendland outperforms the C^4-Wendland in terms of loglikelihood; the C^4-Wendland is too smooth since it possesses four continuous derivatives at the origin, while the C^2-Wendland has just two.

The chordal, circular, and Legendre-Matérn fit both datasets reasonably well in terms of Gaussian residual loglikelihoods, whereas all of the other covariance functions fit poorly to at least one of the datasets. While the Bernoulli has the ability to vary the smoothness, it lacks a range parameter, which limits its flexibility. For these datasets, we are not able to detect any large improvement of the circular Matérn or the Legendre-Matérn over the chordal Matérn; the maximum residual likelihoods are slightly different, but we do not interpret this difference to be practically significant. We consider the comparisons presented here to constitute evidence for the wide-ranging applicability of the chordal Matérn, since the two datasets are quite different in terms of smoothness. Moreover, the chordal Matérn is faster to compute than both the circular Matérn and the Legendre-Matérn; for the satellite data, filling the 2000×2000 covariance matrix for all observations and missing values required 0.85 seconds with the chordal Matérn versus 5.13 seconds and 3.29 seconds with the circular Matérn and Legendre-Matérn (100). All computations were completed on an Intel Core-i7 4770 processor at 3.40 GHz with 32 GB of RAM, running Matlab version R2013a.

5. Discussion

We have studied the mean square differentiability properties of stochastic processes on spheres based on their covariance functions. We prove a theorem characterizing covariance functions for m times mean square differentiable processes on d dimensional spheres. Since it is common to construct processes on spheres by considering restrictions of processes on Euclidean spaces of higher dimension, we prove a corollary stating that the restricted processes retains the differentiability properties of the original process. This result has the important implication that the choral Matérn covariance function retains the full flexibility, in terms of smoothness, of the Matérn covariance function.

The use of chordal distance as the argument in covariance functions has been questioned because great circle distance is the natural distance metric on a sphere. To provide insight into whether the chordal Matérn is an appropriate covariance model for data on spheres, we constructed alternative covariance functions that possess analogous flexibility of smoothness but also respect spherical geometry. The new covariance functions and a number of existing covariance functions were fit to smooth and non-smooth datasets, and their modeling and predictive performance were compared. The chordal Matérn, the circular

Matérn, and the Legendre-Matérn families all contain members that fit well in terms of loglikelihood and provide good predictive performance to both of the datasets. All of the other existing covariance functions achieve suboptimal performance when applied to at least one of the datasets. We conclude that we do not see any evidence that the use of chordal distance introduces any distortions. On the other hand, we do see evidence that considering only classes of covariance functions that are valid on Euclidean spaces and remain valid on spheres with great circle distance does limit our ability to adequately model a wide range of smooth and non-smooth datasets. While we have provided detailed methods for computing the circular Matérn in Appendix S2, the chordal Matérn is still faster to compute, so we recommend the use of the chordal Matérn in applications.

Acknowledgements

This material is based upon work supported by the National Science Foundation's Research Network for Statistical Methods for Atmospheric and Oceanic Sciences (STATMOS) under Grant No. 1107046, and by National Science Foundation Grant No. 1406016.

Appendix A. Proof of Theorem 1

Theorem 3. *If an isotropic process Z on $\mathbb{S}^d$ has covariance function*

$$\psi(\theta) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} f_k \exp(ik\theta),$$

then the following statements are equivalent:

1. *Z is m times mean square differentiable.*
2. $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$.
3. *$\psi^{2m}(0)$ exists and is finite.*

Proof. $1 \Leftrightarrow 2$: Because Z is isotropic, to prove that 1 is equivalent to 2 it suffices to show that $Z_{\mathbb{X}}^{(m)}(\phi(\boldsymbol{x}))$, as defined above, exists for an $\boldsymbol{x}$ in a great circle $\mathbb{X}$ if and only if $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$. Define Y to be a process on the real line such that $Y(\phi(\boldsymbol{x}) + 2\pi k) = Z_{\mathbb{X}}(\phi(\boldsymbol{x}))$ for $k \in \mathbb{Z}$. Stein (1999) defines a process Y on the real line to be mean square differentiable at ϕ if $\lim_{\varepsilon \to 0}(Y(\phi + \varepsilon) - Y(\phi))/\varepsilon$ exists in the Hilbert space of linear combinations of Y, with higher order derivatives defined similarly as above. Therefore, $Z_{\mathbb{X}}$ is m times mean square differentiable at $\boldsymbol{x}$ if and only if Y is at $\phi(\boldsymbol{x})$. Embedding a restriction of Z in a process on the real line allows us to use classical results on mean square differentiable processes on the real line. Due to Bochner's theorem, isotropic covariance functions on the real line can be expressed as

$$\mathrm{Cov}(Y(\phi_1), Y(\phi_2)) = \int_{\mathbb{R}} e^{i\omega(\phi_1 - \phi_2)} dF(\omega),$$

where $F(\omega)$ is called the spectral measure. Stein (1999) shows that if a process on the real line has spectral measure F, then it is m times mean square differentiable if and only if $\int_{\mathbb{R}} \omega^{2m} dF(\omega) < \infty$. The process Y that we constructed has covariance function

$$\mathrm{Cov}(Y(\phi_1), Y(\phi_2)) = \sum_{k=-\infty}^{\infty} e^{ik(\phi_1 - \phi_2)} f_k,$$

Thus, its spectral measure is $F(\omega) = \sum_{k \leq \omega} f_k$, and

$$\int_{\mathbb{R}} \omega^{2m} dF(\omega) = \sum_{k=-\infty}^{\infty} k^{2m} f_k.$$

Hence Y, and therefore Z, is m times mean square differentiable if and only if $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$.

$2 \Leftrightarrow 3$: The finiteness of the $2m$'th moment of a positive finite measure is equivalent to the existence and finiteness of the $2m$'th derivative of its characteristic function (Chung, 2001, Theorem 6.4.1), so $\sum_{k=-\infty}^{\infty} k^{2m} f_k < \infty$ is equivalent to $\psi^{2m}(0)$ existing and finite.

□

References

Anderes, E.B., Stein, M.L., 2008. Estimating deformations of isotropic Gaussian random fields on the plane. Annals of Statistics 36, 719–741.

Banerjee, S., 2005. On geodetic distance computations in spatial modeling. Biometrics 61, 617–625.

Chen, D., Menegatto, V.A., Sun, X., 2003. A necessary and sufficient condition for strictly positive definite functions on spheres. Proceedings of the American Mathematical Society , 2733–2740.

Chung, K.L., 2001. A Course in Probability Theory. 3 ed., Academic Press.

Digital Library of Mathematical Functions, 2012. URL: http://dlmf.nist.gov/. release 1.0.5 of 2012-10-01.

Du, J., Ma, C., Li, Y., 2013. Isotropic variogram matrix functions on spheres. Mathematical Geosciences 45, 341–357.

Gneiting, T., 2013. Strictly and non-strictly positive definite functions on spheres. Bernoulli 19, 1327–1349.

Guttorp, P., Gneiting, T., 2006. Studies in the history of probability and statistics XLIX: On the Matérn correlation family. Biometrika 93, 989–995.

Heaton, M., Katzfuss, M., Berrett, C., Nychka, D., 2014. Constructing valid spatial processes on the sphere using kernel convolutions. Environmetrics 25, 2–15.

Higdon, D., 1998. A process convolution approach to modelling temperatures in the North Atlantic ocean. Environmental and Ecological Statistics 5, 173–190.

Hitczenko, M., Stein, M.L., 2012. Some theory for anisotropic processes on the sphere. Statistical Methodology 9, 211–227.

Huang, C., Zhang, H., Robeson, S.M., 2011. On the validity of commonly used covariance and variogram functions on the sphere. Mathematical Geosciences 43, 721–722.

Jeong, J., Jun, M., 2015. A class of matérn-like covariance functions for smooth processes on a sphere. Spatial Statistics 11, 1–18.

Johnson, W.P., 2002. The curious history of Faà di Bruno's formula. The American Mathematical Monthly 109, 217–234.

Jun, M., Stein, M.L., 2007. An approach to producing space-time covariance functions on spheres. Technometrics 49, 468–479.

Lang, A., Schwab, C., 2013. Isotropic Gaussian random fields on the sphere: regularity, fast simulation, and stochastic partial differential equations. arXiv preprint arXiv:1305.1170 .

Ma, C., 2012. Stationary and isotropic vector random fields on spheres. Mathematical Geosciences 44, 765–778.

Ma, C., 2014. Isotropic covariance matrix functions on all spheres. Mathematical Geosciences , 1–19.

Marinucci, D., Peccati, G., 2011. Random fields on the sphere: representation, limit theorems and cosmological applications. volume 389. Cambridge University Press.

Marinucci, D., Peccati, G., 2013. Mean-square continuity on homogeneous spaces of compact groups. Electronic Communications in Probability 18, 1–10.

Paciorek, C.J., Schervish, M.J., 2006. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics 17, 483–506.

Sampson, P.D., Guttorp, P., 1992. Nonparametric estimation of nonstationary spatial covariance structure. Journal of the American Statistical Association 87, 108–119.

Schoenberg, I.J., 1942. Positive definite functions on spheres. Duke Mathematical Journal 9, 96–108.

Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer.

Terdik, G., 2013. Angular spectra for non-Gaussian isotropic fields. arXiv preprint arXiv:1302.4049 .

Yadrenko, M.I., 1983. Spectral Theory of Random Fields. Optimization Software.

Yaglom, A.M., 1987. Correlation Theory of Stationary and Related Random Functions I. Springer-Verlag.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. Journal of the American Statistical Association 99, 250–261.

Isotropic Covariance Functions on Spheres: Some Properties and Modeling Considerations
Supplementary Material

Joseph Guinness and Montserrat Fuentes

North Carolina State University, Department of Statistics

S1. Proof for closed form of the circular Matérn

When the smoothness parameter ν is a half-integer, the circular Matérn is given by the formula

$$\psi_n(\theta) = \frac{a}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\exp(ik\theta)}{(\alpha^2 + k^2)^n},$$

where $a = 2\alpha \sinh(\alpha\pi)$.

Theorem 2.

$$\psi_n(\theta) = \sum_{k=0}^{n-1} a_{nk} (\alpha(\theta - \pi))^k \mathrm{hyp}^{(k)}(\alpha(\theta - \pi)),$$

where $\mathrm{hyp}^{(k)}(t)$ *is* $\cosh(t)$ *when* k *is even and* $\sinh(t)$ *when* k *is odd. The coefficient* $a_{n,n-1}$ *is given by*

$$a_{n,n-1} = \left[(-2\alpha^2)^{n-1}(n-1)!\right]^{-1}.$$

For $r = 0, \ldots, n-2$ *and* $k = 0, \ldots, n-1$*, define*

$$h_{rk} = \sum_{j=0}^{2r+1} \binom{2r+1}{j} (k)_j \, (\alpha\pi)^{k-j} \, hyp^{(k-j+1)}(\alpha\pi),$$

where $(k)_j$ *is the falling factorial and equals* 1 *if* $j = 0$ *and equals* $(k)(k-1)\cdots(k-j+1)$ *if* $j > 0$*. We define the matrix* $\boldsymbol{H}_{n-1}$ *to be the* $(n-1)\times(n-1)$ *matrix with* $(r+1, k+1)$*th entry* h_{rk}*, and the* $(n-1)\times 1$ *vector* $\boldsymbol{h}_{n-1}$ *to have* $(r+1)$*th entry* $h_{r,n-1}$*. Then the vector of coefficients* $\boldsymbol{a}_n = (a_{n0}, \ldots, a_{n,n-1})'$

is given by the formula

$$\boldsymbol{a}_n = \begin{bmatrix} (a_{n0}, \ldots, a_{n,n-2})' \\ a_{n,n-1} \end{bmatrix} = \begin{bmatrix} -a_{n,n-1}\boldsymbol{H}_{n-1}^{-1}\boldsymbol{h}_{n-1} \\ a_{n,n-1} \end{bmatrix}.$$

Proof: The function $\psi_n(\theta)$ satisfies the following inhomogeneous differential equation with constant coefficients:

$$\sum_{m=0}^{n-1} c_m \psi_n^{(2m)}(\theta) = \psi_1(\theta),$$

with

$$c_m = \binom{n-1}{m}(-1)^m \alpha^{2(n-1-m)}.$$

We know that $\psi_1(\theta) = \cosh(\alpha(\theta - \pi))$ (Gradshteyn and Ryzhik, 2007, Equation 1.445.2). Making the substitution $t = \theta - \pi$, the differential equation has general solution

$$\psi_n(t+\pi) = \sum_{k=0}^{2(n-1)} (b_{k1}(\alpha t)^k e^{\alpha t} + b_{k2}(\alpha t)^k e^{-\alpha t}),$$

which can be solved by the method of undetermined coefficients. Symmetry conditions on $\psi_n(t+\pi)$ around $t = 0$ require that $b_{k1} = b_{k2}$ if k is even, and $b_{k1} = -b_{k2}$ when k is odd, so the general solution can be rewritten as

$$\psi_n(t+\pi) = \sum_{k=0}^{2(n-1)} b_k (\alpha t)^k \mathrm{hyp}^{(k)}(\alpha t).$$

The $(2m)$th derivative of ψ_n is thus given by

$$\psi_n^{(2m)}(t+\pi) = \sum_{k=0}^{2(n-1)} b_k \alpha^{2m} \sum_{j=0}^{2m} \binom{2m}{j} (k)_j (\alpha t)^{k-j} \mathrm{hyp}^{(k-j)}(\alpha t),$$

and the differential equation is

$$\sum_{m=0}^{n-1} c_m \sum_{k=0}^{2(n-1)} b_k \alpha^{2m} \sum_{j=0}^{2m} \binom{2m}{j} (k)_j (\alpha t)^{k-j} \mathrm{hyp}^{(k-j)}(\alpha t) = \cosh(\alpha t). \quad (1)$$

To evaluate the left hand side of (1), we exchange the order of addition to arrive at

$$\sum_{k=0}^{2(n-1)} \sum_{j=0}^{2(n-1)} \sum_{m=\lceil j/2 \rceil}^{n-1} c_m b_k \alpha^{2m} \binom{2m}{j} (k)_j (\alpha t)^{k-j} \text{hyp}^{(k-j)}(\alpha t) = \cosh(\alpha t), \quad (2)$$

where $\lceil \cdot \rceil$ is the ceiling function. To study (2), we proceed by fixing k and summing over j and m. We can ignore the terms for which $j > k$ because in those cases $(k)_j = 0$. For $k = 0, \ldots, 2(n-1)$, we define

$$\begin{aligned}
p_k(t) &:= \sum_{j=0}^{k} \sum_{m=\lceil j/2 \rceil}^{n-1} c_m \alpha^{2m} \binom{2m}{j} (k)_j (\alpha t)^{k-j} \text{hyp}^{(k-j)}(\alpha t) \\
&= \sum_{j=0}^{k} (k)_j (\alpha t)^{k-j} \text{hyp}^{(k-j)}(\alpha t) \sum_{m=\lceil j/2 \rceil}^{n-1} \alpha^{2m} \alpha^{2(n-1-m)} (-1)^m \binom{n-1}{m} \binom{2m}{j} \\
&= \sum_{j=0}^{k} \frac{1}{j!} (k)_j (\alpha t)^{k-j} \text{hyp}^{(k-j)}(\alpha t) \alpha^{2(n-1)} \sum_{m=\lceil j/2 \rceil}^{n-1} (-1)^m \binom{n-1}{m} (2m) \cdots (2m-j+1) \\
&= \sum_{j=0}^{k} \frac{1}{j!} (k)_j (\alpha t)^{k-j} \text{hyp}^{(k-j)}(\alpha t) \alpha^{2(n-1)} \sum_{m=\lceil j/2 \rceil}^{n-1} (-1)^m \binom{n-1}{m} P_j(m),
\end{aligned}$$

where $P_j(m) = (2m)(2m-1)\cdots(2m-j+1)$ is a jth order polynomial in m that equals zero when $m = 0, \ldots, \lceil j/2 \rceil - 1$, so we can allow the sum over m to run from 0 to $n-1$ in

$$p_k(t) = \sum_{j=0}^{k} \frac{1}{j!} (k)_j (\alpha t)^{k-j} \text{hyp}^{(k-j)}(\alpha t) \alpha^{2(n-1)} \sum_{m=0}^{n-1} (-1)^m \binom{n-1}{m} P_j(m).$$

Since P_j is a jth order polynomial, the sum over m is zero when $j < n-1$ (Gradshteyn and Ryzhik, 2007, Equation 0.154.3), and hence $p_k(t)$ is zero when $k < n-1$. When $j = n-1$,

$$\begin{aligned}
\sum_{m=0}^{n-1} (-1)^m \binom{n-1}{m} P_j(m) &= \sum_{m=0}^{n-1} (-1)^m \binom{n-1}{m} \left((2m)^{n-1} + Q_{n-2}(m) \right) \\
&= \sum_{m=0}^{n-1} (-1)^m \binom{n-1}{m} (2m)^{n-1} \\
&= 2^{n-1} (-1)^{n-1} (n-1)! \\
&\neq 0, \qquad (3)
\end{aligned}$$

where the second equality is due to the fact that Q_{n-2} is a polynomial of degree $n-2$, and the third equality follows from Gradshteyn and Ryzhik (2007, Equation 0.154.4). If we rewrite $p_k(t)$ for $k \geq n-1$ as

$$p_k(t) = \sum_{j=n-1}^{k} \frac{1}{j!}(k)_j (\alpha t)^{k-j} \mathrm{hyp}^{(k-j)}(\alpha t) \alpha^{2(n-1)} \sum_{m=0}^{n-1} (-1)^m \binom{n-1}{m} P_j(m), \tag{4}$$

we can now see that the set of functions $p_k(t)$ for $k = n-1, \dots 2(n-1)$ are linearly independent, since $p_k(t)$ can be written as

$$p_k(t) = \sum_{l=0}^{k-(n-1)} w_{kl} (\alpha t)^l \mathrm{hyp}^{(l)}(\alpha t),$$

where w_{kl} are constants for which $w_{k,k-(n-1)} \neq 0$ by (3). The differential equation in (1) is now reduced to

$$\sum_{k=n-1}^{2(n-1)} b_k p_k(t) = \cosh(\alpha t).$$

with the set of functions $p_k(t)$ linearly independent. When $k = n-1$, we have

$$\begin{aligned} p_{n-1}(t) &= b_{n-1} \cosh(\alpha t) \alpha^{2(n-1)} \sum_{m=0}^{n-1} (-1)^m \binom{n-1}{m} \left((2m)^{n-1} + Q_{n-2}(m) \right) \\ &= b_{n-1} \cosh(\alpha t) \alpha^{2(n-1)} 2^{n-1} (-1)^{n-1} (n-1)!, \end{aligned}$$

which is proportional to $\cosh(\alpha t)$. Since the differential equation is set equal to $\cosh(\alpha t)$ and the p_k's are linearly independent for $k \geq n-1$, we have determined that

$$\frac{1}{b_{n-1}} = \left(-2\alpha^2\right)^{n-1} (n-1)!.$$

We now have the equation

$$\sum_{k=n}^{2(n-1)} b_k p_k(t) = 0. \tag{5}$$

Since the p_k's are linearly independent, (5) implies that $b_k = 0$ for all $k > n-1$. However, the other coefficients $b_0, \dots, b_{n-2}$ must be determined by enforcing the initial conditions implied by ψ_n. We know that the odd derivatives of ψ_n up to order $2(n-1)-1$ must be zero when evaluated at 0 and 2π. This gives us $n-1$ initial conditions. The $(2r+1)$th derivative of ψ_n evaluated at $\theta = 2\pi$

(or $t = \pi$) is

$$\psi_n^{(2r+1)}(2\pi) = \sum_{k=0}^{n-1} b_k \alpha^{2r+1} \sum_{j=0}^{2r+1} \binom{2r+1}{j} (\alpha\pi)^{k-j} (k)_j \text{hyp}^{(k-j+1)}(\alpha\pi).$$

Setting this equation equal to zero gives us the $n-1$ equations corresponding to $r = 0, \ldots, n-2$:

$$\sum_{k=0}^{n-2} \left[\sum_{j=0}^{2r+1} \binom{2r+1}{j} (\alpha\pi)^{k-j} (k)_j \text{hyp}^{(k-j+1)}(\alpha\pi) \right] b_k$$
$$= -b_{n-1} \sum_{j=0}^{2r+1} \binom{2r+1}{j} (\alpha\pi)^{n-1-j} (n-1)_j \text{hyp}^{(n-j)}(\alpha\pi).$$

With h_{rk}, $\boldsymbol{H}_{n-1}$, and $\boldsymbol{h}_{n-1}$ defined as above, the solution of this system of equations is

$$(b_0, \ldots, b_{n-2})' = -b_{n-1} \boldsymbol{H}_{n-1}^{-1} \boldsymbol{h}_{n-1}.$$

S2 Approximation for arbitrary smoothness

In most cases, the smoothness of the process is not known *a priori*, so it is desirable to have methods for estimating the smoothness from the data. This generally requires computing exactly or approximating the covariance function with arbitrary values of ν. As far as we know, the circular Matérn does not have a closed form expression in terms of elementary or special functions of θ when ν is not a half-integer. As a result, we are forced to resort to an approximation, but in this case we show that there is a computationally efficient approximation with good theoretical properties.

According to the Poisson summation formula (Zwillinger, 2003), the circular Matérn can always be written as

$$\psi(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \frac{e^{ik\theta}}{(\alpha^2 + k^2)^{\nu+1/2}} = \sum_{n=-\infty}^{\infty} M(\theta + 2\pi n), \tag{1}$$

where M is the continuous Fourier transform of $f(\omega) = (\alpha^2 + \omega^2)^{-\nu-1/2}$, so that $M(\theta)$ is proportional to $\mathcal{K}_\nu(\alpha\theta)(\alpha\theta)^\nu$ with proportionality constant depending on α and ν. Consider a truncation of the right hand sum in (1), obtaining

$$C_N(\theta) = \sum_{n=-N}^{N} M(\theta + 2\pi n).$$

This approximation should be numerically sufficient in most cases when α is not too small, and ν is not too large, i.e. when neither the range nor the smoothness

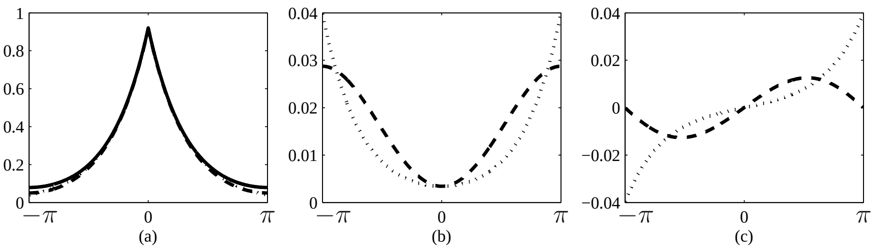

Figure 1: (a) $\psi(\theta)$ with $\nu = 1/2$, $\alpha = 1$ (solid line), along with $\widetilde{\psi}(\theta)$ (dashed line) and $C_0(\theta)$ (dotted line), (b) $R(\theta) = \psi(\theta) - \widetilde{\psi}(\theta)$ (dashed line) and $R_0(\theta) = \psi(\theta) - C_0(\theta)$ (dotted line), (c) $R^{(1)}(\theta) = \psi^{(1)}(\theta) - \widetilde{\psi}^{(1)}(\theta)$ (dashed line) and $R_0^{(1)}(\theta) = \psi^{(1)}(\theta) - C_0^{(1)}(\theta)$ (dotted line).

are too large. For example, when $\nu = 1/2$, M decreases exponentially with rate α.

However, the truncated approximation does not carry with it any theoretical guarantees of positive definiteness or "closeness" to ψ. If we add an even polynomial to $C_N(\theta)$, such theoretical results are possible. To this end, we propose the approximation

$$\widetilde{\psi}(\theta) = C_N(\theta) + p_{2d}(\theta),$$

where $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$ is chosen so that $R(\theta) := \psi(\theta) - \widetilde{\psi}(\theta)$ is $2d$ times continuously differentiable on the unit circle $\mathbb{T}$. Controlling the derivatives of the difference $R(\theta) = \psi(\theta) - \widetilde{\psi}(\theta)$ between a covariance function and an approximation to it is central to proving results relating to both the positive definiteness of the approximation and the extent to which $\widetilde{\psi}$ is a good approximation to ψ, where we define the approximation to be good if the resulting Gaussian measures are equivalent.

Defining $R_N(\theta) := \psi(\theta) - C_N(\theta)$, we rewrite $R(\theta) = R_N(\theta) - p_{2d}(\theta)$. It can be shown (Lemma 2 in Appendix S3), that $R_N(\theta)$ is infinitely continuously differentiable on $[-\pi, \pi]$. However, the derivatives at $-\pi$ and π are not necessarily equal to each other, so R_N is not infinitely continuously differentiable on the unit circle. Specifically, R_N is an even function, so its even derivatives are continous on the unit circle, but its odd derivatives may be discontinuous at one point on the unit circle. However, if we choose p_{2d} so that its odd derivatives at $-\pi$ and π up to order $2d-1$ match those of R_N, then R will be $2d$ times continuously differentiable on the unit circle, since p_{2d} is also an even and infinitely continuously differentiable function on $[-\pi, \pi]$. In Figure 1, we show an example with $d = 1$ and $N = 0$.

In general, in order for the odd derivatives of p_{2d} at $-\pi$ and π to match those of R_N, we set $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$, with $a_{2d} = R_N^{(2d-1)}(\pi)/((2d)!\pi)$, and proceeding recursively, set

$$a_{2(d-j)} = \frac{1}{\pi(2(d-j))!}\left[R_N^{(2(d-j)-1)}(\pi) - \sum_{k=0}^{j-1}\frac{(2(d-k))!}{(2(j-k)+1)!}a_{2(d-k)}\pi^{2(j-k)+1}\right] \tag{2}$$

for $j = 1, \ldots, d-1$. The value of a_0 does not affect the differentiability. The computation of these odd derivatives appears at first to be a daunting task since R_N is an infinite sum. However, when $\theta = \pi$, this sum can be rewritten as

$$R_N(\pi) = M(-\pi(2N+1)) + \sum_{n=1}^{\infty} M(-\pi(2N+1+2n)) + M(\pi(2N+1+2n)).$$

Since M is even, when R_N is differentiated an odd number of times, each term in the sum will be zero, and we can compute any odd derivative of R_N evaluated at π by simply computing the odd derivative of M evaluated at $-\pi(2N+1)$.

The following theorem establishes positive definiteness of the approximate covariance function.

Theorem 3. *If $d > \nu$, there exists an even polynomial $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$ and a finite integer N for which $\widetilde{\psi}(\theta) = C_N(\theta) + p_{2d}(\theta)$ is positive definite on $\mathbb{S}^1$, and if $d > \nu + 1/2$, there exists even polynomial and finite integer N for which $\widetilde{\psi}$ is positive definite on $\mathbb{S}^2$ and $\mathbb{S}^3$.*

Proof. To prove that $\widetilde{\psi}(\theta)$ is positive definite on $\mathbb{S}^1$, it is sufficient to show that $\widetilde{f}_k = \int_{\mathbb{T}} \widetilde{\psi}(\theta)e^{-ik\theta}d\theta > 0$ for all $k \in \mathbb{Z}$, and $\sum_{k\in\mathbb{Z}} \widetilde{f}_k < \infty$. Using $R(\theta) = \psi(\theta) - \widetilde{\psi}(\theta)$, $\widetilde{f}_k$ can be expressed as

$$\widetilde{f}_k = \int_{\mathbb{T}} (\psi(\theta) - R(\theta))e^{-ik\theta}d\theta = f_k - \varepsilon_k,$$

where f_k and ε_k are the Fourier coefficients for ψ and R, respectively. Defining a_{2k} as in (2) for $k = 1, \ldots, d$, choose N large enough so that $\sup_{\theta\in[-\pi,\pi]} R_N^{(j)}(\theta) < (1/2)(\alpha^2+1)^{-\nu-1/2}$ for both $j = 2d$ and $j = 2d-1$, and so that

$$\int_{\mathbb{T}} |R_N(\theta)|^2 d\theta + \int_{\mathbb{T}} |p_{2d}(\theta)|^2 d\theta < f_0^2,$$

which are all possible due to Lemma 2 in Appendix S3. Then $R(\theta) = R_N(\theta) - p_{2d}(\theta)$ is $2d$ times continuously differentiable on $\mathbb{T}$, and $R^{(2d)}$ is differentiable everywhere on $\mathbb{T}$ except at π. Furthermore, $R^{(2d)}(\theta)$ is bounded above by $(\alpha^2 + 1)^{-\nu-1/2}$ because $p_{2d}^{(2d)}(\theta) = R_N^{(2d-1)}(\pi)$, and N was chosen so that $R_N^{(2d-1)}(\theta)$ and $R_N^{(2d)}(\theta)$ were both bounded above by $(1/2)(\alpha^2+1)^{-\nu-1/2}$.

Using Lemma 9.5 in Körner (1989), the differentiability properties of R imply that $|\varepsilon_k| < A|k|^{-2d-1}$ for $k \neq 0$, where $A = (\alpha^2+1)^{-\nu-1/2}$. One can check that $f_k = (\alpha^2+k^2)^{-\nu-1/2} \geq |k|^{-2\nu-1}(\alpha^2+1)^{-\nu-1/2}$ for $k \neq 0$. Therefore, if $d > \nu$, then $|\varepsilon_k| < f_k$ for every $k \neq 0$, and thus $\widetilde{f}_k > 0$ for every $k \neq 0$.

The sum

$$\sum_{k\in\mathbb{Z}} |\varepsilon_k|^2 = \int_{\mathbb{T}} |R(\theta)|^2 < \int_{\mathbb{T}} |R_N(\theta)|^2 d\theta + \int_{\mathbb{T}} |p_{2d}(\theta)|^2 d\theta < f_0^2,$$

which implies that $|\varepsilon_0| < f_0$, so that $|\varepsilon_k| < f_k$ for every k, and thus $\widetilde{f}_k > 0$ for every k. Finally, $|\varepsilon_k| < f_k$ for every k, and $\sum f_k < \infty$, imply that $\sum \widetilde{f}_k = \sum f_k - \varepsilon_k < \infty$. Therefore, $\widetilde{\psi}$ is positive definite on the circle.

To prove that $\widetilde{\psi}$ is positive definite on $\mathbb{S}^2$ and $\mathbb{S}^3$, it is sufficient to show that $\widetilde{f}_k - \widetilde{f}_{k+1} > 0$ for every $k \geq 0$. By the generalized binomial theorem, when $k > \max(1, \alpha^2)$,

$$\begin{aligned} f_k &= k^{-2\nu-1} + c_1 k^{-2\nu-2} + o(k^{-2\nu-2}), \\ f_{k+1} &= (k+1)^{-2\nu-1} + c_2 (k+1)^{-2\nu-2} + o(k^{-2\nu-2}), \\ &= k^{-2\nu-1} + c_3 k^{-2\nu-2} + o(k^{-2\nu-2}), \end{aligned}$$

with f_k strictly monotonically decreasing implying that $c_0 = c_1 - c_3 > 0$. So we have $f_k - f_{k+1} = c_0 k^{-2\nu-2} + d_k$, where $d_k = o(k^{-2\nu-2})$. Choose $\delta > 0$ such that $2\delta < c_0$. Then

$$f_k - f_{k+1} > (c_0 - \delta) k^{-2\nu-2} + d_k,$$

and there exists $k_0 < \infty$ such that $|d_k| < \delta k^{-2\nu-2}$ for all $k > k_0$. Therefore $f_k - f_{k+1} > (c_0 - 2\delta) k^{-2\nu-2}$ for all $k > k_0$, with $c_0 - 2\delta > 0$.

Choose N large enough that $\sup_{\theta\in[-\pi,\pi]} R_N^{(j)}(\theta) < (1/4)(c_0 - 2\delta)$ for $j = 2d$ and $j = 2d - 1$ and so that

$$\int_{\mathbb{T}} |R_N(\theta)|^2 d\theta + \int_{\mathbb{T}} |p_{2d}(\theta)|^2 d\theta < \frac{1}{4} \left(f_k - f_{k+1}\right)^2$$

for every $k \leq k_0$. Then we can bound $|\varepsilon_k| < (1/2)(c_0 - 2\delta)|k|^{-2d-1}$ for $k \neq 0$ as before, and thus $|\varepsilon_k - \varepsilon_{k+1}| < (c_0 - 2\delta) k^{-2d-1}$ for all $k > k_0$. Therefore, if $d > \nu + 1/2$, $|\varepsilon_k - \varepsilon_{k+1}| < f_k - f_{k+1}$ for every $k > k_0$.

Since $\sum_{j\in\mathbb{Z}} |\varepsilon_j|^2 < (1/4)(f_k - f_{k+1})^2$ for every $k \leq k_0$, we have $|\varepsilon_j| < (1/2)(f_k - f_{k+1})$ for every $k \leq k_0$ and for every j. Therefore, $|\varepsilon_k - \varepsilon_{k+1}| < f_k - f_{k+1}$ for every $k \leq k_0$. Thus $\widetilde{f}_k - \widetilde{f}_{k+1} > 0$ for every k.

□

The following theorem asserts that it is possible to construct the approximation so that not only is it positive definite, but it well approximates the true covariance function, in that the two Gaussian measures are equivalent. We denote the Gaussian measure on $\mathbb{S}^1$ with mean 0 and covariance function ψ by $G(0, \psi)$.

Theorem 4. *If $d > \nu + 1/4$, there exists an even polynomial $p_{2d}(\theta) = \sum_{k=0}^{d} a_{2k}\theta^{2k}$ and a finite integer N for which the Gaussian measures $G(0,\psi)$ and $G(0,\widetilde{\psi})$ on $\mathbb{S}^1$ are equivalent.*

Proof. Using results in Stein (1999), the two Gaussian measures $G(0,\psi)$ and $G(0,\widetilde{\psi})$ are equivalent if $f_k = O(\widetilde{f}_k)$, $\widetilde{f}_k = O(f_k)$, and the following sum is finite:

$$\sum_{k\in\mathbb{Z}} \frac{\left(f_k - \widetilde{f}_k\right)^2}{f_k^2} = \sum_{k\in\mathbb{Z}} \frac{\varepsilon_k^2}{f_k^2}.$$

In the proof of Theorem 3, we showed that it is possible to choose N large enough so that $|\varepsilon_k| < A|k|^{-2d-1}$ for all $k \neq 0$ and that $f_k > A|k|^{-2\nu-1}$. Therefore, if $d > \nu + 1/4$, the Fourier coefficients are of the same order, and $\varepsilon_k^2/f_k^2 < |k|^{-(1+\delta)}$ for some $\delta > 0$ and for all $k \neq 0$, so the sum converges. □

We conjecture that it is possible to choose d and N large enough so that the Gaussian measures are equivalent on $\mathbb{S}^2$ and $\mathbb{S}^3$, but a proof of that would involve describing conditions for equivalence of Gaussian measures on higher order spheres, which is beyond the scope of this paper.

S3 Proofs of lemmas

Lemma 1. *For $\alpha > 0$, $\nu, \mu \in \mathbb{R}$, $j \in \mathbb{Z}$, the sequence of functions*

$$S_n(\theta) = \sum_{0<|k|\le n} \operatorname{sgn}(k)^j \mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^\mu$$

is uniformly convergent on $[-\pi, \pi]$.

Proof. Using the Cauchy criterion, our aim is to show that for every $\varepsilon > 0$, there exists N such that for every $n, m \ge N$, $\theta \in [-\pi, \pi]$ implies that $|S_n(\theta) - S_m(\theta)| \le \varepsilon$. First we see that if $n, m \ge N$, then

$$|S_n(\theta) - S_m(\theta)| \le \sum_{|k|>N} \mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^\mu$$

since $\mathcal{K}_\nu(x) > 0$ for $x > 0$ (Digital Library of Mathematical Functions, 2012, 10.37). Using the fact that $\mathcal{K}_\nu(x)$ is decreasing in x (Digital Library of Mathematical Functions, 2012, 10.37, 10.27.3), and that for $\theta \in [-\pi, \pi]$, $2\pi|k| - \pi \le |\theta + 2\pi k| \le 2\pi|k| + \pi$, we obtain $\mathcal{K}_\nu(\alpha|\theta + 2\pi k|) \le \mathcal{K}_\nu(\alpha(2\pi|k| - \pi))$, and $(\alpha|\theta + 2\pi k|)^\mu \le (\alpha(2\pi|k| + \pi)^\mu$ for $\mu \ge 0$. We assume $\mu \ge 0$ because if $\mu < 0$, the summand is eventually bounded by $\mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^0$. Therefore

$$|S_n(\theta) - S_m(\theta)| \le 2 \sum_{k>N} \mathcal{K}_\nu(\alpha(2\pi k - \pi))(\alpha(2\pi k + \pi))^\mu,$$

so if this sum converges, we can always find N such that $|S_n(\theta) - S_m(\theta)| < \varepsilon$. The modified Bessel function of the second kind has the property that $K_\nu(z) \sim (\pi/2)^{1/2} z^{-1/2} e^{-z}$. This implies that there exists a positive constant A and integer M such that for all $k > M$,

$$\mathcal{K}_\nu(\alpha(2\pi k - \pi)) \leq A\sqrt{\frac{\pi}{2}}(\alpha(2\pi k - \pi))^{-1/2} e^{-\alpha(2\pi k - \pi)}.$$

Thus

$$\sum_{k>M} \mathcal{K}_\nu(\alpha(2\pi k - \pi))(\alpha(2\pi k + \pi))^\mu \leq A\sqrt{\frac{\pi}{2}} e^{2\pi\alpha} \sum_{k>M} \frac{(2\pi k + \pi)^{1/2}}{(2\pi k - \pi)^{1/2}} (\alpha(2\pi k + \pi))^{\mu - 1/2} e^{-\alpha(2\pi k + \pi)},$$

which clearly converges. □

Lemma 2. *For every $N, j \in \mathbb{Z}^+$, $R_N^{(j)}(\theta)$ is continuous on $[-\pi, \pi]$, and for every j, $R_N^{(j)}(\theta)$ converges to zero uniformly on $[-\pi, \pi]$ as $N \to \infty$.*

Proof. We first need to show that $R_N(\theta)$ can be repeatedly differentiated term-by-term so that we can write down expressions for $R_N^{(j)}(\theta)$. Formally, we define

$$R_{N,n}(\theta) = c_{\alpha,\nu} \sum_{N<|k|\leq n} \mathcal{K}_\nu(\alpha|\theta + 2\pi k|)(\alpha|\theta + 2\pi k|)^\nu.$$

Following Theorem 7.17 (Rudin, 1796), suppose the following conditions hold for a sequence of functions f_n:

(a) $f_n(\theta)$ is differentiable on $[-\pi, \pi]$ for each n,
(b) $f_n(\theta_0)$ converges for some $\theta_0 \in [-\pi, \pi]$,
(c) $f_n^{(1)}(\theta)$ converges uniformly on $[-\pi, \pi]$.

Then f_n converges uniformly on $[-\pi, \pi]$ to a function f, and $f^{(1)}(\theta) = \lim_{n\to\infty} f_n^{(1)}(\theta)$ for $\theta \in [-\pi, \pi]$, i.e. the limit of the derivatives of a sequence of functions is equal to the derivative of the limit.

We set $f_n(\theta) = R_{N,n}(\theta)$ and check the conditions of the theorem. If $k \neq 0$, $|\theta + 2\pi k| > 0$ for all $\theta \in [-\pi, \pi]$. Since the modified Bessel function of the second kind and polynomials are both differentiable away from 0, it follows that $R_{N,n}(\theta)$ is also differentiable for each n, so (a) holds. According to Lemma 1, $R_{N,n}$ is uniformly convergent on $[-\pi, \pi]$ as $n \to \infty$, so (b) holds as well. The derivative of the modified Bessel function can be expressed as

$$\mathcal{K}_\nu^{(1)}(x) = -\frac{1}{2}\left(\mathcal{K}_{\nu-1}(x) + \mathcal{K}_{\nu+1}(x)\right) \tag{1}$$

(Watson, 1966), so the derivative of $R_{N,n}(\theta)$ is given by

$$R^{(1)}_{N,n}(\theta) = c_{\alpha,\nu} \sum_{N<|k|\leq n} \frac{\text{sgn}(k)\alpha}{2} \left(-\mathcal{K}_{\nu-1}(\alpha|\theta+2\pi k|) - \mathcal{K}_{\nu+1}(\alpha|\theta+2\pi k|)\right)(\alpha|\theta+2\pi k|)^{\nu}$$
$$+\alpha\nu\mathcal{K}_{\nu}(\alpha|\theta+2\pi k|)(\alpha|\theta+2\pi k|)^{\nu-1},$$

which consists of three terms, each of which can be written in the form in Lemma 1. Therefore $R^{(1)}_{N,n}(\theta)$ converges uniformly, so (c) holds and thus $R_N(\theta)$ can be differentiated once term-by-term.

In general, using repeated applications of (1), the j'th derivative of $R_{N,n}$ can be expressed as

$$R^{(j)}_{N,n}(\theta) = c_{\alpha,\nu} \sum_{N<|k|\leq n} \sum_{l=0}^{j} \text{sgn}(k)^j \alpha^j \binom{j}{l} \left[(-2)^{-j+l} \sum_{m=0}^{j-l} \binom{j-l}{m} \mathcal{K}_{\nu-j+l+2m}(\alpha|\theta+2\pi k|)\right] \times$$
$$\left[\frac{\Gamma(\nu+1)}{\Gamma(\nu+1-l)}(\alpha|\theta+2\pi k|)^{\nu-l}\right]. \quad (2)$$

Since the number of terms is finite, we can exchange the order of summation so that $R_{N,n}(\theta)$ can be written as a finite number of sums of the form given in Lemma 1. Proceeding inductively, we assume that R_N can be differentiated j times term-by-term, with derivative given by the limit as $n \to \infty$ of the expression in (2). Then to complete the induction we must show that $R^{(j)}_N$ can be differentiated term-by-term, which amounts to establishing the conditions of Theorem 7.17 with $f_n = R^{(j)}_{N,n}$. Differentiability holds again due to differentiability of polynomials and the modified Bessel function of the second kind. Convergence at a point holds again because of the form of $R^{(j)}_{N,n}(\theta)$ and Lemma 1. Uniform convergence of the derivative also holds because the $(j+1)$'th derivative of $R_{N,n}$ can also be written as a finite number of sums of the form in Lemma 1. An additional consequence of Theorem 7.17 is that $R^{(j)}_{N,n}$ converges uniformly on $[-\pi,\pi]$. Therefore we have shown that $R_N(\theta)$ can be differentiated term-by-term an arbitrary number of times, and the convergence of the j'th derivative (as $n \to \infty$) is uniform on $[-\pi,\pi]$. Continuity of the derivatives follows from the fact that each derivative is differentiable.

The uniform convergence of the derivatives allows us to easily show that $R^{(j)}_N(\theta)$ converges to zero uniformly on $[-\pi,\pi]$ as $N \to \infty$. Suppose that $N > M$. Then we can write $R^{(j)}_N(\theta) = R^{(j)}_M(\theta) - R^{(j)}_{M,N}(\theta)$. We have just shown that $R^{(j)}_{M,N}$ converges uniformly to $R^{(j)}_M$ as $N \to \infty$, which means that for every $\varepsilon > 0$, we can find N_0 such that $N > N_0$ implies $|R^{(j)}_{M,N} - R^{(j)}_M| < \varepsilon$, which in turn implies that $|R^{(j)}_N(\theta)| < \varepsilon$ for every $N > N_0$. □

References

Digital Library of Mathematical Functions, 2012. URL:

http://dlmf.nist.gov/. release 1.0.5 of 2012-10-01.

Gradshteyn, I.S., Ryzhik, I.M., 2007. Table of Integrals, Series, and Products. 7 ed., Elsevier.

Körner, T.W., 1989. Fourier Analysis. Cambridge University Press.

Rudin, W., 1796. Principles of Mathematical Analysis. McGraw Hill.

Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer.

Watson, G.N., 1966. A Treatise on the Theory of Bessel Functions. Cambridge University Press.

Zwillinger, D., 2003. Standard Mathematial Tables and Formulae. 31 ed., Chapman and Hall.